IOAN TODORAN

Răspunsuri posibile

CORELAȚIE ȘI PROGNOZĂ

dacia

Scanned with OKEN Scanner

IOAN TODORAN

RĂSPUNSURI POSIBILE

CORELAŢIE ŞI PROGNOZĂ

EDITURA DACIA, CLUJ—NAPOCA,
1989

Scanned with OKEN Scanner

CUPRINS

Introducere 7

Capitolul I Rolul prognozei în natură şi societate 9

1.1. Problema prognozei cutremurelor de Pămînt 10
1.2. Unele probleme ale activităţii vulcanice 14
1.3. Problema corelaţiei în meteorologie. Timpul probabil 17
1.4. Rolul factorilor cosmici în probleme de prognoză 20
1.5. Importanţa problemelor de prognoză pentru activitatea economică şi socială 27

Capitolul II Noţiuni de teoria semnalului 32

2.1. Semnale aleatoare şi semnale deterministe 32
2.2. Reprezentarea semnalelor: 34
a) Aproximarea semnalului printr-un polinom 34
b) Reprezentarea semnalelor prin funcţii periodice 37

Capitolul III Noţiuni de teoria corelaţiei 52

3.1. Colectivitate statistică 53
3.2. Frecvenţa unui eveniment dat 55
3.3. Valoarea medie şi dispersia unei distribuţii empirice 57
3.4. Corelaţia 58
3.5. Corelaţia liniară. Ecuaţia dreptei de regresie 60
3.6. Determinarea coeficientului de corelaţie 65
3.7. Proprietăţile coeficientului de corelaţie 69
3.8. Corelaţia rangurilor 71
3.9. Corelaţia seriilor dinamice 78
3.10. Corelaţia neliniară. Raportul de corelaţie 95
3.11. Corelaţii multiple şi parţiale 102
3.12. Funcţia de corelaţie 114

Capitolul IV Corelaţie—Cauzalitate—Prognoză 128

4.1. Stabilirea formulelor empirice 129
4.2. Utilizarea metodei celor mai mici pătrate în forma elaborată de Gauss 134
4.3. Metoda polinoamelor ortogonale 136
4.4. Stabilirea formulelor semi-empirice. Metoda corecţiilor diferenţiale . . 148
4.5. Interpolarea şi extrapolarea funcţiilor empirice 151

Capitolul V Utilizarea graficelor pentru probleme de prognoză 158

5.1. Importanţa reprezentării grafice în problemele de corelaţie şi cauzalitate 158
5.2. Formarea punctelor normale. Trasarea curbelor medii 161
5.3. Ajustarea datelor experimentale. Netezirea curbelor empirice 164

Scanned with OKEN Scanner

5.4. Deviația standard a unei curbe empirice 174
5.5. Rolul simetriei curbelor empirice 176
5.6. Proiecte de prognoză bazate pe corelație 179

Capitolul VI **Prognoza fenomenelor periodice** 182
6.1. Caracteristicile principale ale unei funcții periodice 182
6.2. Ameliorarea valorilor perioadelor medii 190
6.3. Formule empirice pentru prognoza fenomenelor periodice 200
6.4. Prognoza fenomenelor periodice cu perioadă variabilă 217
6.5. Prognoza fenomenelor multiperiodice. 218
Bibliografie selectivă 221

Scanned with OKEN Scanner

INTRODUCERE

Legătura dintre om și mediul înconjurător constituie o problemă care, din timpuri străvechi, este în atenția gîndirii umane. Dar modul de a aborda această chestiune a fost diferit în epoci diferite.

Astfel, mitologiile religiilor primitive încercau să exprime, prin acțiunea forțelor supranaturale, influența pe care fenomenele naturale o exercită asupra omului. În acest sens putem aminti o serie de calamități naturale care au fost acceptate cu resemnare de către omenire și împotriva cărora ar fi fost inutil să luptăm, așa cum inutilă ar fi lupta împotriva morții inevitabile.

Așadar, fenomenele naturale ca : erupțiile vulcanice, cutremurele de Pămînt, inundațiile, secetele, perioadele de foamete, epidemiile, taifunurile, curenții violenți mareici, furtunile etc. au fost acceptate ca fiind o manifestare a forțelor supranaturale și, de aceea, o luptă împotriva lor era considerată inutilă și fără sorți de izbîndă. Omul apărea slab și neputincios față de puterea naturii dezlănțuită sub forma fenomenelor catastrofice amintite mai sus.

Dar o dată cu dezvoltarea metodelor de cercetare științifică — teoretică și practică, omul a început să cunoască treptat anumite cauze ale fenomenelor naturale, a început să înțeleagă modul lor de desfășurare și, prin aceasta, a început să prevadă unele manifestări care se vor produce în viitor și să ia măsuri de apărare împotriva lor. Astfel, de exemplu, după descoperirea paratrăsnetului în 1747 de către Benjamin Franklin (1706—1790), s-au luat măsurile necesare pentru apărarea împotriva trăsnetului. În mod similar, după ce Louis Pasteur (1822—1895) a descoperit o serie de microbi, s-au putut lua măsuri de luptă împotriva unor epidemii distrugătoare. În fine, după construirea barajelor pentru hidrocentrale s-a reușit să se reducă considerabil rolul distrugător al inundațiilor și secetelor.

Lupta dintre om și natură nu este ușoară, știința încă nu a creat modele concrete pentru multe fenomene naturale. De aceea, deocamdată, pe baza unui bogat material de date observaționale, putem stabili corelații empirice a căror existență să ne servească la cercetarea unor probleme de prognoză. Încercarea de a da o formă matematică pentru diferite probleme de corelație și utilizarea lor în aplicații practice, constituie subiectul prezentei monografii.

Scanned with OKEN Scanner

Unele dintre cele mai distrugătoare fenomene naturale, a căror prognoză este încă un subiect de discuţie, sînt prezentate, pe scurt, în capitolul I. Este vorba de cutremurele de Pămînt, erupţiile vulcanice şi fenomenele meteorologice, pentru prognoza acestora fiind încă deosebit de interesante unele cercetări statistice bazate pe corelaţii. Tot aici este inclus şi rolul factorilor cosmici de care trebuie să ţinem seama atunci cînd se fac studii de prognoză privind activitatea economică şi socială.

În capitolul II sînt prezentate cîteva noţiuni de teoria semnalului, noţiuni deosebit de importante pentru analiza funcţiilor empirice.

Problemele de corelaţie propriu-zise apar în capitolul III unde sînt prezentate cîteva dintre noţiunile frecvent întîlnite în cercetările de care se face uz în capitolele următoare. Exemplele practice prezentate aici ilustrează modul practic de a cerceta legăturile posibile dintre diferiţi parametri fizici prin care se caracterizează modul de manifestare a unui fenomen natural sau proces social.

În capitolul IV se discută metode pentru stabilirea unor formule empirice necesare la problemele de prognoză. În mod deosebit se insistă asupra utilizării metodei celor mai mici pătrate în forma elaborată de Gauss şi a polinoamelor ortogonale.

Capitolul V conţine noţiuni privind utilizarea graficelor în probleme de prognoză: formarea punctelor medii, netezirea curbelor empirice etc. Tot aici sînt expuse cîteva modele de prognoză bazate pe corelaţie.

În fine, capitolul VI este consacrat funcţiilor empirice periodice şi utilizarea acestora pentru prognoză.

Elaborarea acestui volum reprezintă o continuare a activităţii autorului în domeniul cercetării funcţiilor empirice. A trecut aproximativ un deceniu de la apariţia lucrării „Tratarea matematică a datelor experimentale" (Edit. Acad. RSR, 1976), timp în care autorul a adunat material suplimentar şi a elaborat metode noi privind utilizarea funcţiilor empirice în probleme de prognoză. Iată de ce, în cursul expunerii, se fac mai multe referiri la lucrarea publicată anterior.

În cursul expunerii s-au folosit noţiuni de matematică clasică care pot fi uşor înţelese de către cititorul absolvent de şcoală medie. Am procedat astfel pentru ca materialul expus aici să fie accesibil tuturor acelora care, sub o formă sau alta, se interesează de prognoză, chiar dacă nu au o pregătire deosebită în domeniul matematicii.

Scanned with OKEN Scanner

CAPITOLUL I

ROLUL PROGNOZEI ÎN NATURĂ ȘI SOCIETATE

Din păcate, mai există încă multe fenomene ale căror cauze sînt necunoscute sau știm prea puține lucruri despre ele. În astfel de situații nu se poate vorbi despre o formulare matematică a modului lor de manifestare și, prin aceasta, nu se dispune de o predicție sigură. Totuși, și în astfel de situații trebuie să avem un mijloc de prevedere, trebuie să putem elabora o prognoză, chiar dacă ea va fi bazată pe un grad redus de probabilitate. În acest scop, avînd la dispoziție un număr mare de date observaționale, se pot pune în evidență corelații existente între diferiți parametri fizici, se stabilesc relații empirice corespunzătoare, iar printr-o extrapolare potrivită se prevede forma de *desfășurare probabilă* a fenomenelor cercetate. De-altfel, corelația asigură doar o funcție probabilă și nu o dependență funcțională riguroasă.

În vederea justificării necesității unei teorii cît mai dezvoltate privind problemele de corelație și, în același timp, cît mai accesibile cercetărilor practice, amintim aici cîteva probleme, în studiul cărora metodele corelației pot fi de mare folos. Astfel de probleme se referă la fenomenele care pot fi clasificate în cel puțin patru categorii mari: a) fenomene a căror desfășurare este determinată de agitația existentă în interiorul adînc al planetei noastre, cum sînt cutremurele de Pămînt și erupțiile vulcanice, b) fenomene meteorologice a căror desfășurare are loc în atmosfera terestră, c) factori cosmici mai importanți pentru problemele de prognoză și d) fenomene și evenimente sociale a căror prognoză este deosebit de importantă pentru viața cotidiană.

După cum se știe, a prevedea — a elabora o prognoză — înseamnă a cunoaște și a putea anunța din timp modul probabil de desfășurare a unor evenimente sau fenomene. De aceea, noțiunea de *prognoză* este întîlnită în toate domeniile de activitate umană, deși acest cuvînt este frecvent utilizat în meteorologie — prognoza vremii sau comunicarea timpului probabil în buletinele meteorologice.

De fapt, a vorbi despre prognoză înseamnă a ne interesa de evenimente verificabile și despre care nu știm apriori că se vor produce cu certitudine.

În general, prognoza conține un anumit grad de incertitudine cu privire la apariția unui eveniment probabil, iar incertitudinea res-

Scanned with OKEN Scanner

pectivă poate fi condiționată de diferite cauze pe care nu le cunoaștem destul de bine. Dar oricît de puțin probabilă ar fi o prognoză, ea reprezintă ceva mai mult decît ignoranța.

1.1. Problema prognozei cutremurelor de Pămînt

Din cauza imperfecțiunilor simțurilor noastre, avem impresia că scoarța Pămîntului rămîne nemișcată. Acest fapt este o aparență, o impresie de calm deoarece în realitate scoarța Pămîntului este supusă unor oscilații permanente, adică ea suferă peste tot ușoare „zguduituri". Aceste oscilații neîncetate ale solului — imperceptibile pentru simțurile umane — pot fi înregistrate cu ajutorul unor aparate foarte sensibile numite *seismografe*. Activitatea de „respirație" a Pămîntului se numește *agitație seismică*.

Uneori, din motive încă nu destul de bine cunoscute, în anumite locuri și pe neașteptate, în scoarța terestră se produc o serie de zguduituri și oscilații care se succed și se propagă în toate direcțiile. Astfel de mișcări bruște constituie fenomenul natural cunoscut sub denumirea de *cutremur de Pămînt*. Locul unde încep zguduiturile se numește focar sau *hipocentru* iar proiecția focarului pe suprafața terestră se numește *epicentru*.

Adîncimea focarului unui cutremur se măsoară de-a lungul verticalei corespunzătoare de la epicentru la focar. Această adîncime este de obicei mai mică de 100 km, dar există și cazuri cînd astfel de fenomene se produc la circa 700—800 km sub scoarța terestră. Cutremurele de Pămînt mai puternice produc distrugeri de bunuri materiale și pierderi de vieți omenești pe suprafața terestră, aceste distrugeri fiind numite *efecte macroseismice*.

Cutremurul de la 4 martie 1977 a avut epicentrul în Munții Vrancei, focarul fiind situat la circa 110 km adîncime. Energia eliberată de acesta echivala cu aceea a 100 bombe atomice de tipul celor de la Hiroșima și era de magnitudinea 7,2 pe scara Richter.

Reamintim aici faptul că magnitudinea unui cutremur depinde de energia totală eliberată, iar legătura dintre acestea este exprimată printr-o formulă matematică stabilită de către Ch. Richter în deceniul IV al secolului nostru. De aici avem scara Richter pe care, cutremurele abia sesizabile au magnitudinea zero, pe cînd cele mai violente fenomene de acest fel, cunoscute pînă acum, au avut magnitudinile 8,6—8,8.

Aproape că nu există lună din an în care să nu se înregistreze un cutremur mare, în timp ce cutremurele mici pe care le simt numai seismografele sînt mult mai numeroase. În general există regiuni în care cutremurele apar mai frecvent: Chile, Mexic, Japonia etc. Această situație a impus necesitatea observării atente a activității

Scanned with OKEN Scanner

seismice din regiunile respective precum și încercarea de a prevedea astfel de catastrofe în scopul de a ști cînd și unde urmează a se declanșui astfel de fenomene și, eventual, de a putea evacua populația pentru a reduce la minim efectele dăunătoare.

Pînă în prezent, oamenii sînt incapabili să lupte împotriva cutremurelor de Pămînt sau să influențeze modul lor de desfășurare. Singura apărare împotriva lor este de a examina condițiile de apariție, de a putea ști dinainte cînd urmează a se produce declanșarea lor și de a reuși să se evacueze bunurile materiale și populația din raza de acțiune. De aceea, examinarea acestor fenomene, corelarea lor cu anumiți parametri fizici este foarte importantă. Prognoza activității seismice este de primă necesitate mai ales în anumite regiuni de pe suprafața terestră unde sînt mai frecvente astfel de fenomene.

Foarte probabil că un cutremur își anunță „venirea" cu mult timp înainte, neajunsul principal fiind cuprins în aceea că semnalele respective încă nu sînt cunoscute și înțelese în măsură suficientă pentru a fi utilizate ca elemente de bază în prognoză.

Pentru a rezolva problema prognozei cutremurelor de Pămînt, ar trebui să se dea răspunsuri la următoarele întrebări:

— În ce loc urmează să se producă cutremurul?

— Care va fi intensitatea maximă pe care o va avea?

— La ce dată urmează a se produce cutremurele cele mai puternice?

Referindu-ne la primele două întrebări: prognoza locului și intensitatea cutremurelor, se pot da răspunsuri destul de bune deoarece dispunem de hărți bine întocmite cu privire la raionarea seismică a suprafeței terestre.

În privința celei de a treia întrebări, legată de momentul apariției unui cutremur, cercetările sînt abia la început și nu pot da un răspuns cu o probabilitate mai mare de 98%. Tocmai acesta este motivul pentru care am considerat aici acest gen de probleme în care teoria corelației ar putea aduce un aport substanțial.

Avînd în vedere considerațiile de mai sus, pentru examinarea prognozei cutremurelor de Pămînt, vom deosebi două metode mai importante: Metoda seismologilor chinezi, care ține seamă de apariția diferitelor fenomene premergătoare unui cutremur și metoda care se bazează pe cercetarea undelor seismice.

a) *Metoda seismologilor chinezi* se bazează pe observarea și urmărirea evoluției anumitor fenomene care de obicei apar înaintea izbucnirii unui cutremur. Astfel, pe baza înregistrărilor cutremurelor de Pămînt care s-au produs într-o perioadă de peste 2.800 ani, ale căror caracteristici mai importante au fost descrise în cronicile chineze, se pot întocmi cataloage și hărți izostatice cu o precizie destul de mare pentru scopul urmărit. Din datele cuprinse în aceste cataloage se pot pune în evidență fenomenele anormale care preced de-

clanşarea unui cutremur, se poate stabili dinainte data cînd urmează a se declanşa cutremurul respectiv şi se poate determina mărimea aproximativă a intensităţii lui, precum şi locul în care se va produce. Pe baza unor astfel de constatări se poate face evacuarea ordonată a populaţiei şi a celor mai importante bunuri : animale, inventar mobil etc.

Avînd în vedere faptul că acest mod de prevedere a cutremurelor de Pămînt ar putea fi utilizat şi în alte domenii în care se cere prognoza anumitor fenomene naturale, aducem aici cîteva precizări cu privire la parametrii de interes.

Seismologii chinezi, printr-o propagandă susţinută şi o muncă politică convingătoare, au organizat reţele de observatori voluntari, neprofesionişti, dar care au fost instruiţi din acest punct de vedere.

În modul acesta au fost organizate brigăzi provinciale care observă comportarea anumitor factori :

1) anomaliile electrice şi magnetice,

2) deformarea scoarţei terestre,

3) fluxul de radon în fîntîni (radonul este un element radioactiv alcătuit din mai mulţi izotopi),

4) variaţia undelor seismice,

5) acalmii (răstimp de linişte) seismice care preced cutremurele de Pămînt,

6) variaţii ale nivelului apelor în fîntîni,

7) zgomote acustice premergătoare, şi

8) comportarea animalelor.

Urmărirea parametrilor enumeraţi mai sus este necesară deoarece s-a constatat că între modificarea acestora şi declanşarea unui cutremur există o anumită corelaţie, fapt care este de mare importanţă în prognoza acestor fenomene naturale.

b) *Metoda undelor seismice* se bazează pe studiul variaţiei raportului dintre viteza undelor longitudinale, V_p, şi a celor transversale, V_s. În acest scop, geofizicienii au cercetat corelaţia dintre anumite fenomene seismice secundare şi apariţia seismului propriu-zis. În modul acesta Observatorul seismic din oraşul japonez Kagoshima, într-un anumit interval de timp, a înregistrat 91 de microseisme în loc de media obişnuită corespunzătoare care era de 34 seisme, a dat alarmă şi s-a evacuat oraşul cu încă 18 localităţi din vecinătatea lui, fapt care a făcut să se evite pagubele materiale mari pe care le putea provoca cutremurul care s-a declanşat la cîteva săptămîni după acţiunea de evacuare.

În 1948, seismologul japonez Hayakowa considera că înaintea apariţiei unui cutremur ar putea exista o variaţie a vitezei undelor longitudinale, V_p, dar, din nefericire, datele de care dispunea el erau de slabă calitate astfel că lucrările sale au fost repede uitate.

Scanned with OKEN Scanner

În anul 1960, în regiunea Garm din Uniunea Sovietică (R.S.S. Tadjikă), a fost întreprinsă o vastă cercetare a activității seismice, cercetare care s-a bazat pe o serie de măsurători ale parametrilor : rezistivitatea electrică a terenurilor, nivelul apelor în puțuri, viteza undelor seismice longitudinale și transversale precum și variațiile locale ale cîmpului magnetic. Rezultatele obținute de seismologii sovietici au condus la o confirmare a ideilor menționate mai sus ale lui Hayakowa ; s-a ajuns la concluzia că viteza undelor de șoc a unor cutremure îndepărtate se modifică în valoarea absolută și că în lunile care preced declanșarea unui cutremur puternic raportul V_p/V_s scade atunci cînd ele traversează regiunea examinată. Paralel cu acestea s-a pus în evidență o scădere a rezistivității zonei subterane de falii imediat înaintea producerii cutremurului. De asemenea, s-au observat modificări ale intensității curenților electrici telurici și a cîmpului magnetic, variații care se explică prin faptul că înaintea unui cutremur, din cauza compresiunii, mineralele suferă anumite modificări la nivelul structurii atomice corespunzătoare.

Cercetările efectuate ulterior au dus la concluzii și mai ferme, cu rezultate mult mai complexe. Astfel, înainte de a izbucni cutremurul de la San Fernando (California) în 1971, s-a observat că raportul V_p/V_s scădea pentru un timp oarecare însă, pe urmă, această scădere era urmată de o revenire la normal, iar durata anomaliei acestui raport depindea de magnitudinea cutremului pe care-l prevedea. Pe baza acestor observații s-a putut prevedea, cu precizie destul de bună, locul și data declanșării cutremurului respectiv.

Cercetătorii sovietici au putut prevedea cu trei luni înainte apropierea apariției a două cutremure, în acest scop bazîndu-se pe creșterea observată a concentrației de heliu și radon în anumite puțuri adînci.

Aceste constatări sînt valabile pentru cutremurele al căror focar se găsește la adîncime mică sub suprafața terestră, dar nu rămîn valabile pentru cutremurele cu focare situate la adîncimi de peste 100 km cum sînt cutremurele din Vrancea. Pentru astfel de situații încă nu există metode eficace de prognoză, în schimb există numeroase date observaționale care arată că un cutremur mare necesită un timp îndelungat pentru a se acumula tensiunile corespunzătoare.

În declanșarea unui cutremur puternic se eliberează o mare cantitate de energie și epuizează toate forțele acumulate. De aceea trebuie să treacă mai mulți ani pentru ca un nou cutremur distrugător să se poată declanșa.

Paralel cu adunarea datelor empirice și observarea corelației existente între anumiți parametri ai activității seismice, un rol deosebit încep să-l joace cercetările teoretice în prognoza cutremurelor de Pămînt. În acest scop se iau în considerare o serie de condiții fizice în care se desfășoară activitatea seismică și se ajunge la con-

Scanned with OKEN Scanner

cluzia că viteza undelor longitudinale, V_p, crește într-o rocă uscată, pe cînd într-o rocă saturată variațiile acestei viteze sînt aproape inobservabile. De aici se ajunge la explicația că, în timpul cînd începe să se activeze un anumit cutremur, crește presiunea din interiorul rocilor respective și are loc umplerea cu apă a locurilor goale din roci. În funcție de contracția și dilatarea regiunii respective are loc schimbul de lichid și, prin aceasta, se produce variația vitezelor V_p. Pe baza unor astfel de raționamente, în 1973 a putut fi anunțat un cutremur cu două zile înainte de declanșarea lui în statul New York. De asemenea, intensitatea prevăzută a fost în concordanță cu datele observate. Este de menționat faptul că înaintea declanșării acestui cutremur, raportul V_p/V_s a suferit o variație de circa 13% din valoarea sa obișnuită.

1.2. Unele probleme ale activității vulcanice

În mod obișnuit, prin noțiunea de vulcan se înțelege locul de pe suprafața Pămîntului unde, în mod permanent sau periodic, temperatura este mult mai ridicată decît în punctele învecinate. În aceste locuri, creșterea respectivă a temperaturii este cauzată de apariția magmei (topitură de silicați provenind din părțile adînci ale litosferei) care, datorită proceselor fizice din interiorul Pămîntului, este deplasată spre exterior. Suma tuturor proceselor legate de apariția magmei la suprafața scoarței terestre se numește *vulcanism*.

Fenomenele vulcanice au atras atenția omului încă din cele mai vechi timpuri cînd, în legătură cu ele, au fost create o serie de legende și s-au stabilit diferite mituri.

Spectacolul dezlănțuit de activitatea vulcanică este, fără îndoială, una dintre cele mai mărețe, mai teribile și mai frumoase priveliști pe care ni le oferă natura în diferite împrejurări. Tocmai de aceea, astfel de fenomene au trezit în conștiința omului atît frică și groază cît și admirația, sentimente pe care omul le-a avut întotdeauna față de fenomenele vulcanice.

În zilele noastre este bine cunoscut faptul că vulcanismul este o formă de manifestare a naturii și trebuie să-l integrăm în ansamblul evoluției periferice a planetei noastre; el reprezintă una dintre cele mai caracteristice și mai semnificative trăsături ale fenomenelor geologice.

Noi nu putem și nici nu intenționăm să prezentăm aici o descriere generală a fenomenelor vulcanice și nici a dezastrelor provocate de ele, totuși, pentru problemele care se referă la corelarea fenomenelor naturale, activitatea vulcanică constituie un bun exemplu în care studiile de corelație ar putea aduce contribuții esențiale. Pentru

Scanned with OKEN Scanner

acest motiv prezentăm aici cîteva dintre cele mai semnificative trăsături ale activității vulcanice.

*

* *

După cum se știe, procesele fizice prin care se caracterizează activitatea vulcanică sînt strîns legate de modul de ascensiune a magmelor în litosferă și apariția lor la suprafața terestră. Cu alte cuvinte, în esență, activitatea vulcanică se bazează pe un transfer de substanță și energie calorică din manta sau din zonele adînci ale litosferei spre suprafața Pămîntului.

În timpul marilor erupții vulcanice se pot produce explozii care sînt cauzate de prezența gazelor în magmă. Apoi fluvii de lavă se revarsă peste pantele abrupte ale munților mai cu seamă atunci cînd ascensiunea magmelor are loc prin conducte de formă liniară, chiar dacă secțiunea lor nu este perfect circulară.

Din cauza presiunii interne mari, atunci cînd amestecul de gaze și magmă ajunge la suprafață, se produc ploi de pietre și cenușă, nori incandescenți de gaze toxice care distrug tot ce întîlnesc în drumul lor, provocînd catastrofe îngrozitoare însoțite de moarte și distrugerea unor bunuri materiale. Dar, privită de departe, activitatea unui vulcan se prezintă observatorului ca un spectacol înfricoșător care pune în evidență forța uriașă a naturii în fața căreia, deocamdată, omul apare mic și neputincios.

Din punctul de vedere al frămîntărilor interne, vulcanii se pot asemăna cu oamenii ; fiecare vulcan își are „personalitatea" sa, capriciile sale și evoluează într-un anumit mod în decursul timpului.

Problema prognozei activității vulcanice este foarte greu de abordat deoarece aici apar o serie de obstacole legate direct de natura însăși a fenomenelor respective. Astfel, deoarece vulcanismul își are cauzele principale în străfundul adînc al Pămîntului, este imposibil, deocamdată, să efectuăm observații sau măsurători directe asupra condițiilor respective. De aceea, studiul fenomenelor interioare ale activității vulcanice poate fi asemănat cu activitatea de cercetare a unui medic care prin „ascultare" studiază starea organismului din interiorul unui pacient.

Din cercetarea anumitor stări geochimice și geofizice, precum și din cunoașterea trecutului activității vulcanice a unei anumite regiuni, se pot face totuși pronosticuri pe baza cărora să se evite pagubele mai însemnate și pierderile de vieți omenești. Asupra unor astfel de probleme vom insista în cele ce urmează.

Problema prognozei activității vulcanice trebuie să fie abordată ținînd seama de diversitatea tipurilor de vulcani, de modul lor de manifestare precum și de modul diferit în care are loc ascensiunea materialului magmatic.

Scanned with OKEN Scanner

Știința și tehnica zilelor noastre fac posibilă o supraveghere a vulcanilor, dar, deocamdată, numărul observatoarelor vulcanologice este mult prea mic pentru a dispune de o rețea suficient de mare de date necesare pentru o prognoză demnă de toată încrederea. De asemenea, interpretarea teoretică a datelor observaționale este departe de a fi la nivelul cerințelor unei prognoze; pînă în prezent nu există încă o teorie bine pusă la punct în legătură cu activitatea vulcanică.

În general, „somnul" este starea obișnuită a vulcanilor, iar erupția cu toate consecințele sale constituie un eveniment mult mai rar și de scurtă durată. Prin urmare, prognoza care ne interesează nu este legată de starea permanentă a unui anumit vulcan ci ne interesează „trezirea" lui, erupția și evoluția ulterioară. De aceea, în mod obișnuit, punem întrebarea dacă un anumit vulcan va erupe cu o intensitate maximă și periculoasă? Și dacă da, atunci ar fi deosebit de important să cunoaștem data aproximativă la care se va declanșa erupția respectivă. Pînă în prezent, două metode s-au dovedit a fi mai eficace în privința prognozei erupțiilor vulcanice: I) metoda bazată pe cercetarea propagării undelor seismice și II) metoda care ține seamă de înclinarea părții centrale a vulcanului.

I) Metoda care se bazează pe propagarea undelor seismice are în vedere faptul că activitatea seismică din regiunile vulcanice poate fi influențată în mod diferit de către diferiți vulcani. Astfel, de exemplu, erupția vulcanului Vezuviu din anul 79 î.e.n. a fost precedată de o serie de cutremure care au avut loc pe un interval de timp de cîțiva ani. Dar problema dificilă care se pune aici este aceea de a se putea descifra cutremurele care prevestesc erupțiile vulcanice din mulțimea de cutremure care caracterizează activitatea seismică asupra căreia am insistat în paragraful precedent.

Utilizînd seismografe de diferite sensibilități, echipa Observatorului de vulcanologie din Hawaii a prevăzut „deșteptarea" vulcanului Kilanea (Hawaii) cu aproximativ 24 ore înainte. În acest caz, seisme speciale au fost provocate de ascensiunea magmelor, ceea ce a făcut ca erupția respectivă să poată fi prevăzută cu o precizie destul de bună. Este vorba numai de o prognoză locală unde s-a utilizat o extrapolare a vitezei medii de ascensiune a magmelor.

II) Metoda care se bazează pe cercetarea înclinării părții centrale a vulcanului utilizează un aparat numit *clinograf*. În cazul prognozei erupției vulcanului Kilanea, cele două metode au fost utilizate aproape simultan.

Într-un an, magma bazaltică a vulcanului Kilanea a urcat peste 50 km prin litosferă pentru a atinge suprafața, unde s-a transformat în grandioase revărsări de materie incandescentă care s-a scurs peste panta muntelui respectiv cu o viteză de peste zece metri pe secundă.

Metoda amintită aici nu se aplică în cazuri generale deoarece

Scanned with OKEN Scanner

fiecare vulcan își are capriciile sale, pentru prognoză fiind necesară și cunoașterea altor aspecte. Astfel bunăoară ar fi indicat să se cerceteze prezența unei corelații între diferiți parametri măsurabili ai activității eruptive cum sînt mai cu seamă aceia care se referă la faza gazoasă, care, dealtfel, este elementul mobil al vulcanismului. Prin cercetarea gazelor eruptive se poate indica starea spre care evoluează activitatea unui anumit vulcan.

Prin studii de gravimetrie și magnetometrie se poate determina poziția magmei la anumite adîncimi, iar prin metode electronice se poate determina modificarea conductibilității electrice în straturile mai adînci. De aici se poate desprinde apoi corelația care există între diferiți parametri: fizici, geofizici, chimici și geochimici. Astfel de corelații nu se pot pune în evidență decît numai atunci cînd se produc variații în mărimile respective, adică în cursul derulării fenomenului respectiv — erupția vulcanică. La fel cum geneticienii lucrează efectuînd experiențe pe muște sau bacterii al căror ritm de reproducție este foarte rapid, și nu asupra omului, tot astfel și vulcanologii trebuie să caute legile de dominare a activității vulcanice cercetînd și observînd nu vulcanii stinși ci pe cei activi. Singură activitatea eruptivă și modificările ei rapide pot duce la descoperirea corelației existente între diferiți parametri și faza maximă a erupției.

Dintre corelațiile cele mai importante cunoscute amintim aici pe aceea care există între ascensiunea magmelor premergătoare erupțiilor și temperatura solului. Iradierea calorică a unei regiuni poate fi apreciată cu mare exactitate iar variația acesteia ne poate indica apropierea unei erupții vulcanice. Astfel, de exemplu, erupția vulcanului Etna din 1971 a fost prevăzută cu 15 zile înainte, folosind în acest scop supravegherea încălzirii solului.

Din cercetările efectuate asupra mai multor vulcani s-a constatat că o încălzire pronunțată a solului de pe o regiune relativ mică este un indiciu al apariției unei erupții vulcanice, dar nu orice erupție este însoțită de un astfel de fenomen. Iată de ce stabilirea corelațiilor între diferite tipuri de parametri este necesară pentru a avea o prognoză cît mai sigură. Dintre toate acestea va trebui să căutăm corelația care să aibă o valabilitate cît mai generală pentru cît mai multe tipuri de vulcani sau pentru o anumită structură vulcanică.

1.3. Problema corelației în meteorologie. Timpul probabil

Fenomenele atmosferice și mai cu seamă cele care au loc în troposferă influențează direct desfășurarea vieții noastre economice și sociale, sînt destul de enigmatice încă și sînt greu de prevăzut. Există totuși o serie de încercări de a elabora o prognoză a vremii și asupra acestei probleme vom insista în cele ce urmează.

Scanned with OKEN Scanner

Din nefericire, în prezent există foarte multe fenomene meteorologice a căror natură nu o înțelegem destul de bine deoarece încă nu există o cunoaștere precisă a anumitor procese fizice elementare (separarea sarcinilor electrice, formarea picăturilor de ploaie etc.). De aceea este foarte greu de formulat ecuațiile care să descrie precis modul de trecere a atmosferei de la o stare la alta. Mai bine zis, ecuațiile fundamentale care descriu structura și comportamentul atmosferei la un moment dat sînt de mult cunoscute, dar nu se pot aplica decît la probleme atmosferice foarte limitate și nicidecum la problemele generale ale prevederii timpului. Astfel de ecuații ar putea fi utilizate în probleme de prognoză, dar pentru aceasta este necesară o rezolvare numerică a lor, problema astfel devenind foarte complicată deoarece avem de a face cu ecuațiile complicate ale hidrodinamicii. În acest din urmă caz sînt necesare foarte multe calcule, aici intervin nu numai punctele de pe suprafața Pămîntului ci și puncte situate la diferite înălțimi în care trebuie să cunoaștem valorile mai multor parametri.

În zilele noastre se fac încercări de modelare a proceselor atmosferice, se fac încercări matematice și se utilizează calculatoarele electronice, dar lipsa unor observații destul de precise și bine repartizate în timp ne împiedică să verificăm studiile elaborate teoretic. În orice caz, meteorologia reprezintă un domeniu vast de activitate, dar în care mai este mult de lucru.

Problema prognozei în meteorologie este mult complicată și prin faptul că aici intervin mase mari de aer a căror stare fizică este influențată de mediul înconjurător ; păturile superioare ale atmosferei sînt expuse la bombardamentul particulelor cosmice, iar cele de jos se lovesc de formele neregulate ale suprafeței terestre, aici intervin forțele de vîscozitate care cauzează amestecul turbulent al aerului. De aceea, anumite sisteme de mase de aer care se formează și pe care le-am putea modela matematic își pierd repede identitatea și provoacă mari perturbații în elaborarea unei prognoze. Tot în cadrul acestor probleme mai trebuie să ținem seamă de prezența lanțurilor de munți precum și de influența termică a oceanelor care au mare rol în durata de existență a sistemelor de mase de aer.

Avînd în vedere faptul că sistemele mari de mase de aer sînt mai stabile decît cele mici, mersul vremii se poate prevedea cu mai mare siguranță atunci cînd este vorba de caractere generale — prevederea în linii mari, — dar dacă ne referim la mase mici — la o prognoză detaliată, problema devine mai greu de rezolvat, iar rezultatele devin mai puțin precise. Tocmai de aceea, previziunea pe termen lung va putea indica doar valorile medii ale unor parametri atmosferici : temperatura medie, proprietățile atmosferice, presiunea barometrică etc. Dar rezultatele demne de încredere se pot obține

Scanned with OKEN Scanner

numai pentru intervale scurte de timp: 3—4 zile și se așteaptă ca într-un viitor apropiat să se ajungă la durata de 4—10 zile.

Meteorologia modernă dispune de informații privind parametrii atmosferici: presiunea și umiditatea aerului, temperatura, direcția și puterea vîntului etc. Astfel de date sînt adunate la un număr mare de stații meteorologice răspîndite pe întreaga suprafață terestră sau sînt culese de aparatele de la bordul unor nave cosmice. Pe baza acestor date se alcătuiesc așa-numitele hărți sinoptice pe care sînt marcate modificările zilnice ale vremii și, în modul acesta, se poate prevedea cu oarecare aproximație starea vremii într-un viitor apropiat. Astfel de hărți dau rezultate destul de bune pentru o perioadă de 24 ore pentru continent și 48 ore pe oceane unde există mai puțini factori perturbatori.

Cînd este vorba de o prevedere pe o durată mai îndelungată, sînt necesare date suplimentare cu privire la circulația aerului din regiunile polare unde se găsesc mai multe stații care furnizează datele necesare.

Din practica observațională s-a ajuns la concluzia că unele modificări în desfășurarea fenomenelor meteorologice ar putea fi prevăzute și după modul de comportare a anumitor viețuitoare, cum ar fi de exemplu agitația și migrația păsărilor sau neliniștea de care sînt cuprinse anumite animale. Astfel de observații ne-ar permite să stabilim o corelație și, prin aceasta, o prognoză cu privire la anumite fenomene meteorologice. De asemenea, o prognoză mai bună s-ar putea face dacă s-ar ține seama, într-un mod mai riguros, de corelația care există între defsășurarea fenomenelor meteorologice și factorii cosmici legați de activitatea solară.

În plus, s-a constatat că vremea suferă o mare instabilitate în epoca echinoțiului de primăvară și că există o scădere a temperaturii pe la mijlocul lunii mai, fapt care se explică prin ruperea unor ghețari din banchiza polară, ghețari care coboară spre ocean.

După cum se știe, prognoza meteorologică se bazează pe constatarea practică cu privire la schimbarea climei în funcție de deplasarea maselor de aer. Astfel, din observațiile efectuate zi de zi se poate face o extrapolare pentru ziua următoare, în acest sens fiind utilizate o serie mai mare de date observaționale. Dar, necunoscînd destul de bine problemele teoretice ale fenomenelor respective, în multe cazuri, prognoza bazată numai pe extrapolare duce la o prevedere eronată și, așteptînd o zi senină, ne putem pomeni cu o zi cu atmosferă instabilă sau chiar cu furtuni puternice.

Procedeele moderne de prevedere a vremii — prognoza meteorologică — se bazează pe o serie de date obținute cu ajutorul sateliților artificiali ai Pămîntului. În acest mod, meteorologii dispun de o imagine mult mai cuprinzătoare cu privire la starea fizică a atmosferei terestre precum și la nebulozitățile existente în ea. Prin urmare,

Scanned with OKEN Scanner

pentru a îmbunătăți prognoza meteorologică se impune din ce în ce mai mult necesitatea de a corela fenomenele meteorologice și activitatea solară, fapt care cere meteorologilor ca la elaborarea unei prognoze să utilizeze o tehnică în care se ține seamă de aceste corelații. Acest lucru s-a și început deja, pe scară mondială; pentru anii 1976—79 a fost elaborat planul *Veghii Meteorologice* în care s-a prevăzut o rețea vastă de mijloace tehnice — stații terestre și navale precum și sateliți artificiali. În acest timp s-au obținut date prețioase cu privire la starea fizică a atmosferei atît la suprafața Pămîntului cît și la diferite altitudini.

Am amintit aici doar cîteva probleme privind dificultățile care se ivesc atunci cînd se elaborează o prognoză meteorologică. Alături de cutremurele de Pămînt și erupțiile vulcanice, fenomenele meteorologice sînt prea puțin cunoscute, studiul acestora necesită încă eforturi susținute, succesul unor astfel de cercetări fiind condiționat de datele observaționale obținute prin diferite metode și stabilirea de corelații între diferiți parametri. În general, condițiile fizice din atmosfera terestră trebuie privite în corelație cu anumiți factori cosmici care pot să acționeze asupra stării fizice a fenomenelor meteorologice, atmosfera terestră nefiind un corp izolat în spațiu. Influența mediului exterior joacă un rol deosebit aici, dar asupra acestei probleme vom reveni în paragraful următor.

1.4. Rolul factorilor cosmici în probleme de prognoză

Cercetările efectuate în ultimul timp au arătat, de exemplu, că atunci cînd ne referim la prognoza vremii — timpul probabil — este absolut incorect a lua în considerare numai cauzele terestre. Singură cunoașterea desfășurării fenomenelor naturale terestre nu este suficientă pentru prevederea vremii pe o perioadă mai îndelungată. Aici trebuie luați în considerare și o serie de factori cosmici care au o influență directă asupra Pămîntului și asupra atmosferei sale.

O problemă deosebit de importantă în studiile de prognoză o are ciclicitatea observată a unor fenomene terestre. În legătură cu cauzele acesteia există păreri diferite: unii susțin că ciclicitatea observată este cauzată de factori cosmici — factori extratereștri, alții consideră că la baza ciclicității se găsesc cauze pur terestre, iar alții sînt de părere că ciclicitatea respectivă ar fi de natură întîmplătoare. Prin urmare, din enumerarea acestor păreri contradictorii se vede că aici lipsesc studiile sistematice ale acestor fenomene, lipsesc considerațiile teoretice care să explice în mod analitic caracteristicile ciclice rezultate din anumite date observaționale.

Indiferent de natura care stă la baza periodicității anumitor fenomene terestre, punerea acestora în corelație cu anumiți factori

Scanned with OKEN Scanner

cosmici este deosebit de importantă pentru problemele de prognoză. Pe Pământ există periodicități observate care nu se pot contesta: succesiunea zilelor și nopților ca o consecință a mișcării de rotație a Pământului, succesiunea anotimpurilor ca o consecință a mișcării Pământului pe orbita sa circumsolară, dezvoltarea plantelor în funcție de energia primită de la Soare, grosimea inelelor de creștere anuală a copacilor în funcție de perioadele după care se succed anii secetoși sau ploioși etc.

În paragraful 1.3 am arătat că prognoza vremii pe termen lung se bazează pe calcule cu privire la desfășurarea proceselor hidrodinamice care au loc în atmosfera terestră. Aici se pune accent pe încălzirea straturilor superioare ale atmosferei și transmiterea acesteia la straturile inferioare. Dar această încălzire se datorește Soarelui, iar modul de desfășurare a acestui fenomen a devenit cunoscut abia de curînd în urma cercetărilor efectuate cu ajutorul sateliților artificiali ai Pământului.

În mod analog se prezintă și problemele privind climatologia, hidrologia, glacialogia, seismologia, aeronomia, biologia, medicina și multe altele. Studiile acestora trebuie să fie bazate pe regularitățile multianuale observate: revărsarea rîurilor, existența recoltelor bogate, boli epidemice etc.

După cum se știe, energia tuturor fenomenelor terestre provine din micro și macrocosmos. Tocmai de aceea, pentru cercetarea prognozei acestora, este nevoie de cunoașterea condițiilor cosmice care joacă un rol determinant în dinamica respectivă.

Din factorii cosmici mai importanți în problemele care ne interesează aici amintim activitatea solară și razele cosmice.

În cadrul activității solare, printre cele mai remarcabile fenomene amintim: petele solare, protuberanțele, erupțiile cromosferice și vîntul solar.

— *Petele solare* reprezintă cel mai cunoscut mod de manifestare a activității solare. Ele au fost observate din timpuri străvechi, iar prin lunetă au fost privite prima dată de către Galileo Galilei în anul 1610.

O pată solară este o formațiune de culoare „întunecată" care apare printre granulele fotosferice. La început ea apare ca un por care apoi se dezvoltă și poate să se mențină o săptămînă sau chiar mai mult.

Din studii statistice s-a constatat că activitatea petelor solare, adică numărul acestora și aria ocupată de ele, variază ciclic cu o perioadă de circa 11 ani. Această „periodicitate" este cunoscută sub denumirea de *ciclul activității solare* și este foarte importantă deoarece odată cu modificarea petelor solare au loc și alte variații în modul de manifestare a activității solare și este în strînsă corelație cu desfășurarea unor fenomene de pe suprafața terestră.

Scanned with OKEN Scanner

Activitatea petelor solare este periodică și, prin urmare, intensitatea acesteia va avea valori maxime și minime, de aceste valori extreme fiind legat și modul de manifestare a anumitor fenomene terestre. Iată de ce, atunci cînd vorbim de prognoză, în mod direct sau indirect sîntem nevoiți a ne referi la activitatea solară și, evident, la factorii cosmici.

— *Protuberanțele* sînt formațiuni de materie gazoasă care se pot observa sub forma unor țîșnituri în atmosfera solară. Ele au forma unor șuvoaie de apă care țîșnesc din fîntînile arteziene sau pot să apară ca niște limbi de foc care se înalță deasupra cromosferei, sînt mai puțin strălucitoare decît fotosfera și se pot observa numai în timpul eclipselor totale de Soare sau cu un aparat special numit spectroscop de protuberanțe.

— *Erupțiile cromosferice* sînt un fel de explozii foarte rapide care au loc în cromosfera solară. Ele se caracterizează prin apariția unor „pete" foarte luminoase, evoluează rapid și au o durată de circa 20—40 minute. În acest timp, în cromosferă apare o proeminență mică (înălțimea maximă este de circa 30.000 km) în straturile superioare ale cromosferei. Atît durata cît și înălțimea mică ne ajută ca să nu confundăm aceste proeminențe cu protuberanțele.

În general, erupțiile cromosferice sînt legate de grupurile de pete solare. S-a observat că o dată cu apariția erupțiilor cromosferice crește brusc cîmpul magnetic al petelor solare situate în vecinătatea punctelor în care se produc erupțiile respective. De asemenea, erupțiile cromosferice sînt însoțite de perturbații în ionosfera terestră precum și de o creștere a intensității razelor cosmice înregistrată pe suprafața Pămîntului.

Erupțiile cromosferice se manifestă atît prin radiația electromagnetică din domeniul ultraviolet al spectrului cît și prin radiația corpusculară care constă din particule (corpuscule) ca atomi, ioni și electroni, care, prin erupție, sînt aruncați din Soare cu viteze de ordinul miilor de kilometri pe secundă. Toate aceste particule constituie un „front" care se deplasează de la Soare în toate direcțiile spațiului cosmic, unele dintre ele ajungînd la suprafața Pămîntului.

Erupțiile cromosferice, ca și protuberanțele, se pot observa cu ajutorul spectroscopului de protuberanțe în domeniul spectral definit de linia H_α a hidrogenului, observațiile fiind efectuate de către astronomi. Dar, în același timp, o erupție cromosferică își trădează existența și prin modul în care ea influențează anumite fenomene terestre. Astfel, o erupție cromosferică cauzează o întrerupere brutală a comunicațiilor radio pe unde scurte, mărește intensitatea paraziților atmosferici pe unde lungi și perturbă cîmpul magnetic terestru. Toate acestea se cunosc sub denumirea de *perturbații ionosferice*.

Scanned with OKEN Scanner

Influența unei erupții asupra fenomenelor terestre s-a pus în evidență prin stabilirea unei corelații între perturbațiile ionosferice și momentele la care au fost observate erupțiile respective.

O erupție cromosferică poate elibera o cantitate de energie de $10^{16} - 10^{17}$ kWh (kilowatore) care se propagă în spațiu, iar în atmosfera terestră ajunge sub formă de radiații ultraviolete, raze X și chiar raze γ. Aceste radiații deplasează spre sol straturile ionizate ale ionosferei, fapt care se simte prin întreruperea comunicațiilor radio normale.

Roiurile de particule solare cu energii mari sînt cauza aurorelor polare în care s-au pus în evidență prezența elementelor: azot și oxigen care ar fi componentele obișnuite ale aerului precum și prezența hidrogenului ai cărui atomi trebuie să fie de origine solară.

Este interesant faptul că furtunile magnetice, perturbațiile radio și aurorele polare sînt în corelație strînsă cu numărul de pete solare.

— *Vîntul solar*. Observațiile au arătat că de la Soare, în toate direcțiile, se mișcă curenți de particule cu viteze de ordinul zecilor de kilometri pe secundă, iar pe măsură ce distanța de la Soare crește, crește și viteza particulelor respective. În vecinătatea orbitei terestre viteza corespunzătoare atinge mărimea de cîteva sute de kilometri pe secundă. Această accelerare se explică prin diferența de presiune care există în plasma din vecinătatea Soarelui și aceea din spațiul interstelar. După propunerea astronomului american Parker, acești curenți de corpuscule se numesc *vînt solar*.

Fenomenele amintite mai sus în legătură cu activitatea solară au o influență considerabilă asupra atmosferei terestre și asupra modului de comportare a acesteia. Asupra interacțiunii dintre fenomenele solare și condițiile determinate de acestea pe suprafața Pămîntului există o serie de cercetări teoretice și practice. Această interacțiune este cunoscută sub denumirea de *relații Soare—Pămînt*.

În prezent dispunem de foarte multe date observaționale obținute în domeniul spectral de la lungimile de unde scurte (radiații X și γ) și pînă la lungimile de undă lungi (radiația radio). Cu toate acestea, încă nu există o teorie completă care să poată răspunde la o serie de întrebări: Cum apar petele solare? De ce este ciclică activitatea solară? În ce măsură planetele influențează activitatea solară? etc.

Meteorologia modernă trebuie să țină seama de activitatea solară. Există cercetări care au pus în evidență existența unei corelații între căderile de precipitații și activitatea solară. Perioada precipitațiilor atmosferice este egală cu dublul ciclului activității solare, maximul de ploi corespunde la un ciclu, iar minimul corespunde altui ciclu al activității solare.

În prezent se încearcă a se stabili o legătură între circulația materiei solare (rotația diferitelor straturi, formarea petelor etc.) și acțiunea momentului unghiular al planetelor gigante: Jupiter, Sa-

Scanned with OKEN Scanner

turn, Uranus și Neptun. În acest sens se poate aminti faptul că s-a stabilit deja o corelație între poziția planetei Jupiter și maximele de activitate solară, aceste maxime avînd loc atunci cînd planeta Jupiter se află în punctul de intersecție al orbitei sale cu planul ecuatorial galactic. Din această corelație a rezultat că maximul activității solare depinde de mișcarea sistemului solar prin Galaxie. Prin urmare, factorii galactici influențează modul de desfășurare a activității solare și, prin aceasta, în mod indirect au o influență asupra fenomenelor terestre. De aceea, în cadrul elaborării unei prognoze, va trebui să ținem seamă de cît mai mulți factori cosmici, putînd distinge în acest sens *prognoza pe termen lung* și *prognoza pe termen scurt*.

Prognoza pe termen lung este legată de cercetarea ciclurilor lung periodice ale activității solare, în timp ce prognoza pe termen scurt este legată de examinarea ciclurilor cu periodă scurtă cum sînt, de exemplu, erupțiile cromosferice.

Dacă ne interesează prognoza pe termen de ordinul secolelor, atunci condițiile astrofizice ale Soarelui nu se modifică esențial. De asemenea, nici mișcarea sistemului solar prin Galaxie nu influențează acest mod de prognoză a activității solare, iar pozițiile planetelor uriașe față de ecuatorul galactic se pot cunoaște dinainte. Prin urmare, sistemul solar îl putem considera izolat, iar condițiile care modifică activitatea solară trebuie căutate în interiorul acestuia. În acest sens se ia în considerare modul de distribuție a planetelor în raport cu planul eclipticii și configurația pe care ele o formează în jurul Soarelui.

Din datele observaționale distribuite pe un interval de 200 ani, s-a constatat că modificarea distribuției planetelor mari în jurul Soarelui duce la schimbarea poziției centrului de greutate corespunzător. Din această cauză se schimbă și momentul unghiular al planetelor, iar Soarele suferă variații observabile. Se presupune că, în acest mod, pe suprafața Soarelui apar forțe perturbatoare care favorizează creșterea numărului de pete solare. Ca exemplu se poate arăta că atunci cînd două planete se află pe o dreaptă care trece prin Soare și sînt situate în direcții opuse, forțele care perturbă sînt minime și minim este numărul de pete solare. Cînd planetele respective sînt situate coliniar cu Soarele și de aceeași parte a Soarelui, perturbația este maximă și maximă va fi și activitatea solară.

Cele două coliniarități amintite mai sus depind de perioadele planetelor considerate și, prin aceasta, se poate explica o periodicitate observată în activitatea solară.

Dacă considerăm patru planete: Jupiter, Saturn, Uranus și Neptun, atunci va exista o perioadă de 178 ani în care există 9 conjuncții (planetele fiind situate pe „dreapta" care trece prin Soare și în direcție opusă față de Soare) ale lui Saturn și Jupiter, 4 conjuncții ale lui Uranus și Neptun și o conjuncție a lui Uranus și Neptun. Dealtfel,

Scanned with OKEN Scanner

o serie de cercetări au arătat deja că în analiza activității solare se poate stabili și o periodicitate de 178 ani. În mod analog s-au stabilit cicluri de 500 și 1900 ani, cicluri de care trebuie să se țină seama atunci cînd se elaborează prognoza pe termen lung.

Să ne referim acum la legătura dintre poziția relativă a planetelor față de Soare și prognoza pe termen scurt — ciclurile de cîteva luni sau ani. În acest caz este deosebit de importantă forța mareică a planetelor care acționează asupra suprafeței solare.

Sub acțiunea forțelor de atracție a unei planete asupra suprafeței solare, în straturile fotosferice apare o undă mareică mică, înaltă de circa 1 mm de la fiecare din cele patru planete: Mercur, Venus, Pămînt și Jupiter. Din aceste considerații a fost exclusă planeta Marte deoarece este prea puțin masivă și este situată departe de Soare, iar celelalte planete, care se găsesc dincolo de Jupiter, sînt prea puțin importante în astfel de situații.

În cazul amintit aici, unda mareică rezultantă va fi o funcție de modul în care sînt distribuite planetele respective în jurul Soarelui, ea va avea mai multe valori maxime și minime, iar din analiza lor se vor determina mai multe periodicități.

Pe de altă parte, mediul interplanetar omogen se mișcă cu o viteză supersonică și ajunge la planete, asteroizi sau curenți de meteori de care se lovește și, prin aceasta, apare o undă de șoc ce se mișcă în spațiu odată cu corpul solid corespunzător. De unda de șoc este legată perturbarea parametrilor fizici ai mediului interplanetar, se schimbă densitatea particulelor din mediul respectiv, iar intensitatea cîmpului magnetic suferă variații. În legătură cu aceasta s-a observat deja că numărul de pete solare este în corelație cu procesele furtunoase care perturbă mediul interplanetar.

După modul cum sînt distribuite planetele în jurul Soarelui, undele de șoc ale mai multor planete pot interfera și se pot forma anumite coridoare prin care razele cosmice pot pătrunde cu mai mare ușurință pînă la atmosfera terestră, și cauza o serie de perturbații neprevăzute prin alte considerente. În modul acesta, dintr-o serie de lucrări s-a ajuns la concluzia că atunci cînd Pămîntul și una dintre planetele: Mercur, Venus, Marte și Jupiter sînt coliniare și situate în direcții opuse față de Soare, se observă o reducere a activității geomagnetice care se înregistrează la stațiile terestre.

Astfel de coridoare pot apare și atunci cînd Pămîntul se află în conjuncție cu unele roiuri de meteori sau vreo planetă mică, fapt care ne atrage atenția asupra necesității unei considerații mai riguroase a factorilor cosmici și influenței lor asupra prognozei derulării unor fenomene terestre.

Din punct de vedere observațional, în nenumărate rînduri s-a atras atenția asupra corelației existente între curenții de meteori și mersul vremii (înnourări și precipitații).

Scanned with OKEN Scanner

După cum se știe, cele mai puternice unde de șoc apar în jurul planetelor care au un cîmp magnetic puternic: Pămîntul, Jupiter și probabil Saturn, de aceea poziția acestora poate avea un rol destul de mare în problemele de prognoză.

Evident, în astfel de situații nu se poate exclude Luna, satelitul natural al Pămîntului, care, deși nu are atmosferă și nici cîmp magnetic, provoacă perturbații destul de importante în mediul interplanetar. Aceste perturbații au fost puse în evidență prin detectarea unor corelații ce există între pozițiile Lunii și determinările efectuate de navele cosmice asupra intensității cîmpului magnetic existent în mediul interplanetar. Se consideră că aceste corelații sînt determinate de modul de pătrundere a razelor cosmice pînă la atmosfera terestră unde provoacă perturbații în circulația straturilor atmosferice. Numai în acest mod s-ar putea explica multiplele legături observate între pozițiile Lunii și mersul vremii.

Am amintit mai sus despre acele coridoare care se formează prin suprapunerea mai multor unde de șoc și prin care razele cosmice pot ajunge pînă la atmosfera terestră. De aceea, considerăm necesar să aducem aici încă o mică completare cu privire la modul în care protonii de energie înaltă și probabil particulele α (nucleele de heliu), ar putea să acționeze asupra circulației atmosferei și, prin aceasta, se pune problema cu privire la transformarea energiei potențiale a atmosferei terestre în energie cinetică.

În urma unor cercetări efectuate asupra condițiilor fizice care există în spațiul cosmic, s-a ajuns la concluzia că de acolo, spre atmosfera terestră, vin protoni, particule α și nuclee ale atomilor mai grei. Dar, avînd în vedere faptul că aceste particule pot fi de origine galactică sau proveni de la anumite erupții cromosferice, în funcție de activitatea solară, spre Pămînt poate să vină un flux mai mic sau mai mare de particule a căror energie este de ordinul a $3 \cdot 10^8 - 3 \cdot 10^9$ electronvolți.

După trecerea prin magnetosferă, fluxurile de protoni pot fi condensate de sute de ori și pot stimula trecerea energiei potențiale a atmosferei în energie cinetică care să contribuie la circulația meridiană, și, prin aceasta, devine un suport pentru formarea vînturilor. Ca un exemplu de acest fel poate fi amintit cazul generării de energie cinetică care s-a produs în troposferă în luna ianuarie 1963. În acest timp s-a observat o creștere considerabilă a temperaturii în stratosferă. În a doua și a treia decadă a lunii respective s-a observat o intensificare a activității solare.

La 15 și 26 ianuarie 1963, peste meridianul central al Soarelui au trecut două pete mari. La 14 ianuarie s-a înregistrat o furtună magnetică (perturbații magnetice) și a crescut intensitatea curenților de particule în centura de radiație intensă care se scurgeau spre Pămînt.

Scanned with OKEN Scanner

Din considerațiile prezentate mai sus deducem faptul că între mersul vremii și spațiul cosmic există o corelație puternică și, prin aceasta, între mersul vremii, activitatea solară și proprietățile fizice ale mediului interplanetar există o corelație importantă.

Pentru prognoza pe termen lung trebuie să luăm în considerare și particularitățile cîmpului magnetic interplanetar și fluxurile de particule cosmice care vin spre Pămînt. Prin urmare, factorii cosmici sînt de mare importanță în prognoză și, de aceea, nesocotirea lor poate duce la erori considerabile.

În legătură cu cele amintite mai sus se poate face următoarea obiecție: dacă activitatea solară acționează asupra mersului vremii, atunci de ce într-o regiune vremea este complet diferit de aceea din alte regiuni? La această observație, însă, trebuie să subliniem faptul că activitatea din atmosfera terestră se modifică diferit în locuri diferite, din cauze pe care noi încă nu le cunoaștem destul de bine. Legăturile Soare—Pămînt sînt foarte complicate, sînt chiar mai complicate decît s-ar părea la prima vedere. Tocmai de aceea, în prezent, studiul corelației dintre anumiți parametri observați poate fi de mare interes pentru activitatea teoretică și practică.

1.5. Importanța problemelor de prognoză pentru activitatea economică și socială

Energia care este radiată de Soare se propagă în spațiu, o parte din ea ajunge la atmosfera terestră, pune în mișcare masele de aer, are loc schimbul de căldură dintre regiunile calde și reci, se evaporă apa de pe suprafața Pămîntului, cad ploi și zăpezi, curg rîurile etc. Cu alte cuvinte, energia solară stă la baza tuturor fenomenelor atmosferice cum sînt precipitațiile, nebulozitatea, temperatura și presiunea aerului etc. Prin intermediul energiei solare se creează un mediu de care depinde însăși existența vieții pe planeta noastră.

Pornind de la aceste considerații, în cele ce urmează vom prezenta cîteva dintre aspectele mai caracteristice ale prognozei și rolul ei în activitatea economică și socială.

După cum se știe, în cadrul vieții economice de o deosebită însemnătate sînt problemele privind inundațiile și seceta, fenomene care sînt direct legate de circulația atmosferică și, prin aceasta, sînt supuse influenței activității solare. Consecințele economice care decurg de aici sînt cunoscute, cunoscut fiind și modul în care acestea se simt în toate sectoarele economice.

Dintr-o serie de cercetări statistice s-a constatat că și debitul de apă al rîurilor este corelat cu activitatea solară. În legătură cu aceasta trebuie subliniat faptul că, în anumite regiuni de pe suprafața terestră, debitul de apă al rîurilor variază după anumite perioade

Scanned with OKEN Scanner

bine determinate, pe cînd în alte regiuni există alte perioade sau există o diferență de fază. În acest caz trebuie luate în considerare și particularitățile locale, ceea ce înseamnă că alături de activitatea solară mai intervin și factori locali, iar interferența lor trebuie privită cu multă atenție. Se pare că în unele regiuni există o periodicitate de 6—7 ani după care se produce revărsarea rîurilor și au loc inundații, în alte regiuni perioadele respective putînd fi de 2—3 sau 4—5 ani. Unele dintre aceste periodicități pot fi puse în corelație cu numărul de pete solare. Ar fi deosebit de interesant dacă astfel de corelații s-ar putea lega și de unele fenomene terestre locale și apoi să se separe contribuția lor de contribuția petelor solare. În acest sens teoria corelației (vezi capitolul III) poate fi de mare folos.

În privința secetelor, se poate afirma că pentru o serie de activități economice care sînt strîns legate de agricultură, prognoza poate fi foarte importantă. Este adevărat că datorită mijloacelor tehnicii actuale unele calamități se pot evita prin îndiguiri, baraje și sisteme de irigații, totuși există multe cazuri în care daunele provocate rămîn colosale.

Există o serie de lucrări statistice în care este pusă în evidență dinamica secetelor și corelarea acestora cu producția scăzută la cereale și mai ales la grîu și porumb. Aceste fenomene sînt de asemenea corelate cu unele probleme privind activitatea solară și mai ales cu perturbațiile magnetice.

Deoarece secetele și anii ploioși se repetă în general după o perioadă de 5—6 ani, s-a pus în evidență faptul că recoltele slabe apar atît la maximul cît și la minimul activității solare.

După alți cercetători se pare că anii secetoși și „umezi" alternează și după o perioadă de 30—40 ani. Există și variații care se repetă după 80—90 ani; variații care sînt legate de activitatea solară și pot fi utilizate pentru prognoza climei pentru următoarele două—trei decenii.

Tot pe cale statistică s-a pus în evidență existența unei corelații dintre iernile foarte reci (mai ales în partea răsăriteană a Europei) și o serie de perturbații geomagnetice. Există o periodicitate a recoltelor bogate de grîu care în mare parte este corelată cu numărul de pete solare, periodicitate care însă este afectată și de alte fenomene naturale și sociale cum sînt războaiele sau politica agrară a anumitor țări, diferite transformări sociale etc. Prin urmare, o dată cu dezvoltarea societății se simt influențele sociale care joacă un rol „perturbator" în stabilirea anumitor corelații.

Un criteriu practic pentru punerea în evidență a succesiunii anilor secetoși și „umezi" îl constituie inelele de creștere anuală a copacilor, fapt care duce la următoarele concluzii:

— Anii secetoși și umezi se succed într-o ordine bine determinată.

Scanned with OKEN Scanner

— Anii secetoşi sînt precedaţi, în mare parte, de slăbirea generală a precipitaţiilor atmosferice, deci ei nu vin pe neaşteptate.

— Minimul precipitaţiilor se repetă după 9 sau în unele cazuri după 18 ani.

Cu toate cele menţionate mai sus, între activitatea solară şi ritmul de creştere a copacilor nu există o dependenţă liniară deoarece condiţiile terestre determină multe „perturbaţii" în grosimea inelelor de creştere anuală a copacilor.

Exemple de corelaţie între activitatea solară şi lumea vie se pot aduce foarte multe, unele chiar din lumea animală cum ar fi corelaţia dintre preţul blănurilor de vulpi argintii şi aurorele polare, diferite cicluri observate la peşti sau insecte etc. Şi în literatura zoologică se pot găsi cicluri cu perioadele de ordinul a 4 şi 10 ani.

Periodicităţi se pot pune în evidenţă şi la stabilirea unor modificări ciclice în cazul de îmbolnăvire a oamenilor: gripă, holeră, variolă etc., aceste periodicităţi fiind mult „perturbate" de măsurile sanitare care se iau dinainte. Tocmai de aceea, cunoscînd corelaţia dintre izbucnirile unor epidemii şi alţi factori care ar putea fi observaţi mai dinainte, se pot lua măsuri de prevenire a efectelor dăunătoare. Prin urmare, din acest punct de vedere, medicina şi biologia ar trebui să dispună de informaţii mult mai bogate cu privire la studiul corelaţiilor şi cunoaşterea unor factori legaţi de spaţiul cosmic. Ar trebui să se dispună de date observate pe o mare fracţiune din suprafaţa terestră, date care să fie apoi prelucrate prin mijloacele moderne de calcul.

În lumea vie se cunosc diferite manifestări ciclice, ciclurile cu perioade mai mari de 27—28 zile sînt cauzate de mişcările aparente ale Soarelui şi Lunii.

Într-o serie de lucrări se dau tabele comparative din care se deduce că există corelaţii între anumite procese biologice şi activitatea solară, variaţia ciclică a epidemiilor putînd fi condiţionată de unele oscilaţii ale intensităţii activităţii solare. S-a constatat că există anumite perioade de timp în care mortalitatea printre anumiţi oameni bolnavi creşte simţitor. Deseori se produc decese numeroase în condiţii în care s-ar părea că nu există cauze vizibile care să le producă pe neaşteptate. La autopsia morţilor respectivi se constată că au existat anumite perturbaţii anatomice în muşchiul inimii şi la vasele sanguine, ceea ce arată că organismul respectiv a avut o anumită „pregătire" înainte de a se produce catastrofa.

Uneori se înregistrează „epidemii" de infarcte sau comoţii cerebrale. Astfel, de exemplu, numai în ultimele cinci zile ale lunii ianuarie 1968, la o clinică sovietică, au fost înregistrate atîtea infarcte, cîte se produceau de obicei într-o lună întreagă. Aceste catastrofe sînt puse în corelaţie cu faptul că la acel 28 ianuarie, cînd frecvenţa infarctelor era maximă, pe discul solar a apărut o pată mare. Deşi

Scanned with OKEN Scanner

nu se știe exact care este mecanismul prin care petele solare determină „explozia" de infarcte, corelația amintită este foarte importantă și ne atrage atenția asupra unor chestiuni la care ar trbeui să ne gîndim mai mult.

Din examinarea radiației radio a Soarelui pentru zilele de 27 și 28 ianuarie, 1968, s-a constatat prezența unor intensități maxime care au fost găsite a fi în corelație cu numărul de infarcte. Prin urmare, dacă se determină astfel de corelații, prin observații radio s-ar putea detecta timpul cînd urmează a se produce salturi în intensitatea radiației respective și, prin aceasta, s-ar putea ști cînd să se aplice tratamente preventive persoanelor suspecte de a fi influențate de astfel de factori cosmici.

Mai amintim aici și faptul că există încercări de a stabili corelații între poziția pe care au avut-o corpurile cerești la momentul nașterii unei persoane și destinul său.

Într-adevăr, pînă în prezent nimeni nu a dovedit faptul că Soarele, Luna, planetele, mediul interplanetar, stelele, novele, supernovele, mediul interstelar, galaxiile, quasarii etc. nu au influență asupra personalității umane. Dar tot atît de adevărat este și faptul că încă nimeni nu a descris procesele fizice posibile prin care factorii cosmici acționează asupra organismului nostru. Iată de ce, credem noi, problema corelației dintre factorii cosmici și viitorul noului născut trebuie să fie extrem de complicată. Prognoza bazată pe astfel de corelații, deocamdată, are o probabilitate foarte redusă.

Este adevărat că și în prezent există persoane care mai cred în destinul astral. Dar această credință nu poate fi susținută și nici distrusă prin simple afirmații. Aici este nevoie de o cercetare susținută a factorilor cosmici, domeniu în care se fac investiții considerabile.

Noi nu putem insista aici mai mult asupra acestei probleme, vom reveni asupra ei în paragraful 5.6.

Dezvoltarea impetuoasă a științei și tehnicii care se produce în zilele noastre duce în mod direct la necesitatea elaborării unor planuri de dezvoltare economică și socială în care să fie cuprinse liniile directoare ale noului ritm de dezvoltare. Sînt necesare planuri minuțios întocmite care devin legi pentru sectoarele respective. Dar pentru ca aceste planuri, pe termene mai scurte sau mai lungi, să-și atingă scopul, trebuie ca ele să fie întocmite pe o bază științifică. În acest sens se analizează rezultatele realizărilor precedente, se examinează baza materială disponibilă și, în modul acesta, se pot stabili liniile de dezvoltare — se poate elabora un nou plan în care sînt prevăzute o serie de realizări.

Prevederile științifice în cadrul elaborării unor planuri de dezvoltare rămîn tot prevederi deoarece orice activitate umană este supusă influenței mai multor factori dintre care unii acționează în

Scanned with OKEN Scanner

mod aleator (probabilistic). De aceea, orientarea cercetării științifice spre examinarea și studierea unor astfel de parametri are o importanță considerabilă. Astfel, de exemplu, prevederea științifică a secetelor sau inundațiilor are o valoare economică deosebită, începînd cu agricultura și pînă la cele mai complexe centrale industriale.

În cadrul problemelor economice, pe lîngă prognoza vremii, trebuie să mai cunoaștem și modul de desfășurare a unor fenomene terestre pe care se bazează prognoza unei economii naționale. Astfel, de exemplu, în unele probleme de transporturi pe anumite rîuri ne interesează prognoza debitului de apă al rîurilor respective pentru un interval de timp cît mai mare. O astfel de problemă va avea implicații deosebit de importante în cadrul elaborării unor planuri economice pentru anii următori.

În cadrul anumitor ramuri ale industriei sînt numeroase materii prime a căror disponibilitate depinde și ea de mersul vremii și, prin aceasta, dependența de o serie de factori extratereștri de care trebuie ținut seamă la elaborarea planurilor economice.

Prognoza nu este lipsită de interes nici atunci cînd este vorba de asistența medicală, probleme de învățămînt, probleme de apărare națională etc. Cu un cuvînt, pentru a ști să conduci, printre altele, trebuie să știi să prevezi și să acționezi în sensul potrivit și la momentul potrivit.

Problemele de corelație și prognoză pot conduce la o mai bună planificare a producției bunurilor de consum și la stabilirea unor prețuri echitabile. În modul acesta, pe lîngă asigurarea venitului planificat al întreprinderii respective, trebuie să se satisfacă cererea de pe piață și, în același timp, să se evite stocurile de materiale greu vandabile.

Planurile economice se elaborează pentru un interval de timp în funcție de tematica la care se referă. La noi se cunosc planurile cincinale, dar în afară de acestea mai există planuri de 1, 2, 3 etc. ani. Există planuri de perspectivă care sînt elaborate pe mai mulți ani.

Un plan economic este o lege care prevede ordinea și modul de desfășurare a activității „gospodărești". De aceea, pentru a putea întocmi un plan bun, este necesar să dispunem de: prognoza privind creșterea populației, prognoza resurselor naturale, prognoza dezvoltării tehnico-științifice, prognoza activității sociale și economice etc. Toate aceste prognoze „individuale" se influențează reciproc, fapt ce trebuie avut în vedere atunci cînd se elaborează un plan economic. De toate acestea trebuie să se țină seama atunci cînd se alcătuiește prognoza generală pentru dezvoltarea economiei naționale. Planul de dezvoltare generală trebuie să țină seamă de planurile și prognozele pe ramuri și invers. Nici relațiile internaționale nu trebuie neglijate atunci cînd ne referim la prognoza dezvoltării generale a economiei naționale.

Scanned with OKEN Scanner

CAPITOLUL II

NOȚIUNI DE TEORIA SEMNALULUI

Din considerațiile prezentate în capitolul precedent rezultă că problema prognozei este deosebit de importantă, mai ales atunci cînd ne referim la desfășurarea sau declanșarea anumitor fenomene și procese sociale pentru care nu dispunem de modele matematice prin intermediul cărora să efectuăm o extrapolare exactă. În astfel de cazuri trebuie să cunoaștem modul în care fenomenul respectiv s-a manifestat în ultimul timp, să deducem o „legitate" a acestuia și, prin aceasta, să putem efectua o extrapolare cît de cît probabilă — să dispunem de o prognoză cît mai demnă de încredere. Dar modul de manifestare a unui fenomen sau proces îl putem afla numai prin intermediul semnalelor pe care ni le trimite și pe care noi sîntem capabili să le recepționăm, să le descifrăm și să le interpretăm. Prin urmare, problemele de prognoză sînt strîns legate de anumite caracteristici ale semnalelor recepționate, fapt pentru care, în capitolul prezent expunem cîteva elemente de teoria semnalului.

2.1. Semnale aleatoare și semnale deterministe

După cum se știe, informațiile pe care le avem în legătură cu anumite obiecte sau fenomene ne parvin o dată cu transmiterea unei anumite cantități de energie, transmitere ce se poate realiza pe diferite căi: prin propagarea undelor electromagnetice, propagarea undelor acustice etc. Unda prin intermediul căreia ne parvine informația dorită se numește *semnal* iar unda care este inutilă și care deranjează calitatea recepției se numește *perturbație*.

În cazul cînd este vorba de o undă utilă, dar pe care nu o cunoaștem dinainte, avem de-a face cu un *semnal aleator* (întîmplător), modul de manifestare al acestuia putînd fi prevăzut doar cu un anumit grad de probabilitate. Prin urmare, un semnal aleator poate fi considerat ca un proces stochastic, caracterizat prin mai multe variabile aleatoare, care se desfășoară indefinit în timp și sînt supuse anumitor *legi probabilistice*.

Dacă un anumit semnal are caracterul aleator mai puțin pronunțat, adică dacă imprevizibilul intervine numai în mod parametric,

Scanned with OKEN Scanner

avem de-a face cu un *semnal determinist*. Cu alte cuvinte, semnalele deterministe sînt cazuri particulare ale semnalelor aleatoare, adică un semnal aleator devine semnal determinist atunci cînd semnalul respectiv se poate cunoaște prin determinarea unui număr finit de parametri.

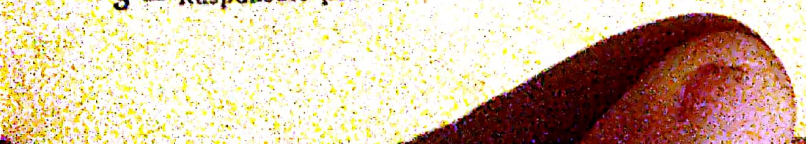
Dacă ne referim la modul de manifestare al unui fenomen, la început am putea avea impresia că avem de-a face numai cu semnale aleatoare, adică s-ar putea să nu dispunem de caracteristicile semnalului respectiv numai prin determinarea unui număr finit de parametri. Observațiile ulterioare, însă, ar putea pune în evidență o „legitate" a unui anumit semnal, ar putea arăta că semnalul aleator se poate caracteriza printr-un număr finit de parametri ; cu alte cuvinte un semnal aleator ar putea deveni semnal determinist. Semnalele deterministe aparțin proceselor deterministe, reprezintă niște idealizări matematice și sînt indispensabile pentru înțelegerea semnalelor aleatoare.

Deoarece semnalele deterministe pot fi exprimate printr-un număr finit de parametri și printr-un număr finit de determinări, modul lor de comportare în timp poate fi prevăzut și, prin aceasta, se rezolvă o problemă de prognoză. În astfel de cazuri, pentru semnalele deterministe se va putea stabili o dependență funcțională — o funcție de timp. Despre astfel de semnale și procese va fi vorba în cadrul acestei monografii.

După modul de înregistrare a unui semnal, putem distinge semnale *continue* și semnale *discrete*. Astfel, intensitatea luminii radiate de către o stea este un proces continuu, dar înregistrarea acesteia poate fi continuă sau discretă. Este vorba de o înregistrare continuă — semnal continuu — cînd, prin intermediul unei celule fotoelectrice, pe o bandă de hîrtie care se derulează uniform în timp, se obține o curbă continuă cu oscilații mai mult sau mai puțin pronunțate. Dar dacă dorim ca înregistrarea respectivă să fie legată de o anumită sursă etalon — o stea de comparație, atunci trebuie ca cele două surse să fie înregistrate succesiv. În acest din urmă caz semnalul va avea un caracter discret, va fi definit printr-o serie de puncte printre care se poate trasa o curbă continuă. Evident, continuitatea curbei respective este o ipoteză pe care o admitem și care rămîne valabilă atîta timp cît nu avem motive să admitem prezența unor discontinuități.

După cum am mai amintit, semnalul nedorit se numește perturbație și denaturează semnalul „adevărat" — semnalul dorit. De aceea, în timpul înregistrării semnalelor dorite, se caută ca influența perturbației să fie minimă sau, dacă este posibil, să fie complet înlăturată, această problemă fiind specifică fiecărui domeniu de cercetare.

După modul în care intervin, perturbațiile se pot împărți în două mari categorii : perturbații aditive și perturbații multiplicative.



Scanned with OKEN Scanner

Dacă $s(t)$ este semnalul dorit, neperturbat, la momentul t și $n(t)$ reprezintă perturbația corespunzătoare, atunci semnalul perturbat și recepționat, $r(t)$, va fi definit prin formula

$$r(t) = s(t) + n(t) \tag{2.1}$$

pentru cazul perturbațiilor aditive, și

$$r(t) = s(t) \cdot n(t) \tag{2.2}$$

pentru perturbațiile multiplicative.

Majoritatea perturbațiilor care intervin în aplicațiile practice sînt de forma (2.1), perturbațiile multiplicative apărînd atunci cînd unul sau mai mulți parametri ai înregistratorului variază în timp.

Dacă $n(t)$ variază în timp, însă foarte încet în comparație cu $s(t)$, atunci perturbația respectivă se numește *feding*.

La receptoarele cu perturbații multiplicative pot să apară și perturbații aditive, iar suprapunerea celor două feluri de perturbații poate duce la probleme foarte complicate.

2.2. Reprezentarea semnalelor

În problemele de prognoză trebuie să stabilim o relație matematică sau o reprezentare grafică prin intermediul căreia să putem efectua o extrapolare prin care se va realiza prognoza. Asupra acestui gen de probleme vom insista în capitolele următoare unde vom examina amănunțit diferite situații care se pot ivi în aplicațiile practice. Aici vom prezenta cazul general astfel încît să avem un schelet de bază și o privire unitară asupra modului de a pune problema și să putem trece cu ușurință de la o formă la alta.

a) *Aproximarea semnalului printr-un polinom*

Dacă $y(x)$ este informația pe care o primim de la un proces aleator la momentul $t = x$, atunci, pentru problemele de prognoză, vom căuta să stabilim o relație care să exprime dependența informației de timp. În cel mai simplu caz, neglijînd rolul perturbațiilor, primele înregistrări ar putea indica o dependență liniară de forma

$$y = a_0 + a_1 x \tag{2.3}$$

Observații mai complete, însă, pot să indice prezența unei relații de forma

$$y = a_0 + a_1 x + a_2 x^2 \tag{2.4}$$

sau o formă generală

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m + \ldots \tag{2.5}$$

Scanned with OKEN Scanner

în care $a_i (i = \overline{1, m})$ sînt mărimi constante pe care trebuie să le determinăm folosindu-ne numai de valorile y_k înregistrate la momentele $t_k = x_k$ $(k = \overline{1, n})$, cu condiția ca numărul ecuațiilor din care rezultă necunoscutele a_i să fie mai mare sau cel puțin egal cu numărul necunoscutelor $(n \geq m)$.

Pentru determinarea necunoscutelor a_i și a gradului m al polinomului respectiv se face uz de principiul celor mai mici pătrate. În acest scop se utilizează metoda lui Gauss (vezi I. Todoran, 1967, capitolul II) sau sub forma stabilită de Cebîșev (1955) cu luarea în considerare a polinoamelor ortogonale.

În cazul cînd utilizăm metoda celor mai mici pătrate, coeficienții a_i se determină astfel încît cu ei să putem calcula mărimile y_k pentru fiecare $t_k = x_k$. În modul acesta, notînd cu $(y_k)_c$ ordonata calculată, metoda celor mai mici pătrate se bazează pe faptul că suma pătratelor diferențelor dintre valoarea observată $(y_k)_{observ.}$ și valoarea calculată $(y_k)_{calc}$ să fie minimă, adică

$$S = \sum_{k=1}^{n} [(y_k)_{observ} - (y_k)_{calc}]^2 = \sum_{k=1}^{n} (0-C)_k^2 = \text{minim} \qquad (2.6)$$

În cazul cînd pentru determinarea coeficienților a_i ne servim de proprietățile polinoamelor ortogonale vom considera că un anumit sistem infinit de funcții reale

$$p_i(x)_{i \in N} \qquad (2.7)$$

este ortogonal pe segmentul $[a,b]$, $(a < b)$, dacă

$$\int_a^b p_n(x) p_m(x) \cdot dx = 0 \text{ pentru } n \neq m \qquad (2.8)$$

$n = 1, 2, 3, \ldots ; \quad m = 1, 2, 3, \ldots$

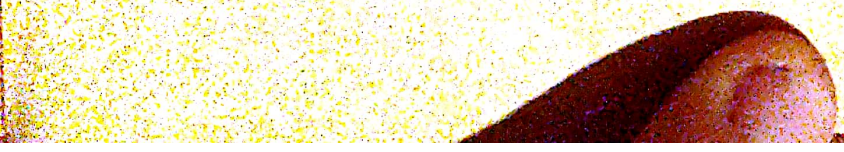
și

$$\int_a^b p_n^2(x)\, dx \neq 0 \text{ pentru } n = m \qquad (2.9)$$

Condiția (2.8) exprimă faptul că funcțiile sistemului (2.7) sînt ortogonale două cîte două, pe cînd condiția (2.9) ne arată că nici una din funcțiile sistemului (2.7) nu este identic nulă.

Sistemul de funcții (2.7) se numește sistem normat dacă

$$\int_a^b p_n^2(x)\, dx = 1 \qquad (2.10)$$



Orice sistem ortogonal se poate norma, fapt care înseamnă că întotdeauna se pot găsi niște constante $\mu_0, \mu_1, \ldots, \mu_n$ astfel încît sistemul

$$\mu_0 p_0(x), \ \mu_1 p_1(x), \ \ldots, \ \mu_n p_n(x),$$

care este și el ortogonal, să fie normat.

Într-adevăr, dacă notăm

$$\int_a^b p_n^2(x)\,dx = \nu_n^2, \qquad \mu_n^2 = \frac{1}{\nu_n^2}$$

atunci

$$\int_a^b \mu_n^2 p_n^2(x)dx = \frac{1}{\nu_n^2}\int_a^b p_n^2(x)dx = 1 \tag{2.11}$$

Mărimea ν_n se numește *norma* funcției $p_n(x)$ și se notează prin simbolul $||p_n||$, adică

$$||p_n|| = \sqrt{\int_a^b p_n^2(x)\,dx}. \tag{2.12}$$

Dacă sistemul (2.7) este normat, atunci

$$||p_n(x)|| = 1 \tag{2.13}$$

Polinoamele ortogonale sînt deosebit de importante în teoria aproximării funcțiilor. Asupra modului de utilizare concretă în astfel de probleme vom reveni în capitolul IV.

În formulele (2.3) — (2.5) am presupus că semnalul se poate reprezenta printr-un polinom care însă, în anumite cazuri, poate deveni incomod pentru problemele de prognoză, mai ales atunci cînd gradul m este destul de mare. În astfel de situații este uneori mult mai avantajos să utilizăm o funcție putere, o funcție exponențială sau să considerăm o funcție periodică mai mult sau mai puțin complicată.

Și în cazul cînd datele experimentale le aproximăm prin forme mai complicate, prin transformări relativ simple, acestea pot fi aduse la forme convenabile pentru aplicațiile practice. Astfel, de exemplu, expresia

$$y = a \cdot b^x$$

poate fi înlocuită prin forma liniară (în planul X, Y)

$$Y = A + BX$$

Scanned with OKEN Scanner

unde am notat

$$Y = \log y; \quad A = \log a, \quad B = \log b, \quad x = X.$$

Transformări asemănătoare se pot utiliza și pentru expresiile:

$$y = a \cdot x^n \Rightarrow Y = A + nX$$

$$y = a \cdot e^{bx} \Rightarrow Y = A + C \cdot x \quad (C = b \cdot \log e)$$

b) *Reprezentarea semnalelor prin funcții periodice*

În cadrul problemelor de prognoză, în mod deosebit ne interesează recepționarea anumitor semnale, precum și modul lor de comportare în timp. De aceea, cîmpul de unde generat de un seism, de erupția unui vulcan etc trebuie să fie considerat atît ca funcție de coordonatele spațiale cît și de timp. Prin urmare, un astfel de cîmp va trebui să fie reprezentat printr-o funcție de forma $f[x(t), y(t), z(t)]$, unde x, y și z sînt coordonatele spațiale care sînt funcții de timpul t. Printre problemele care cer o astfel de abordare vom aminti pe cele ridicate de seismologie, oceanologie, gravimetrie, geomagnetism etc, iar tratarea acestora se bucură de un mare succes atunci cînd se face uz de metodele analizei armonice, asupra căreia vom insista puțin în rîndurile de mai jos.

Evident, utilizarea unei funcții complicate este dificilă și, uneori, precizia datelor observaționale (date empirice) nefiind destul de mare, utilizarea unei funcții prea complicate poate duce la rezultate formale. De aceea, în practică este mult mai convenabil să considerăm cazurile în care una sau cîteva din mărimile variabile sînt menținute constante. Astfel de cazuri ar putea fi de exemplu:

— $f(t)$ care să reprezinte intensitatea unui cutremur înregistrată la o anumită stație care are coordonatele x, y și z (longitudinea, latitudinea și altitudinea) constante.

— $f(x, y, z)$ ar putea reprezenta un cîmp de unde staționar, adică semnalul să fie același în decursul timpului. Astfel de semnale provin de la fenomene care nu sînt variabile în timp, acestea fiind de mică importanță în problemele de prognoză.

— $f(x, t)$ poate să reprezinte un semnal înregistrat la mai multe stații care sînt amplasate pe o anumită direcție (x = variabil).

Printre semnalele cele mai importante pentru problemele de prognoză se află, așa cum s-a mai amintit, semnalele care se pot reprezenta printr-o funcție periodică. În acest caz, funcția $f(x)$, definită pentru toate valorile lui x, se numește *funcție periodică* dacă există o mărime constantă $P \neq 0$ pentru care să putem scrie

$$f(x) = f(x + nP) \tag{2.14}$$

oricare ar fi x. Aici n reprezintă un număr întreg iar constanta P se numește *perioada* funcției $f(x)$.

Scanned with OKEN Scanner

Din examinarea formulei (2.14) rezultă că suma, diferența, produsul și cîtul unor funcții de perioadă P vor fi întotdeauna funcții de aceeași perioadă, fapt care este de mare importanță atunci cînd semnalul recepționat este alcătuit din mai multe semnale elementare.

Dacă construim graficul unei funcții periodice $y = f(x)$ pe un ciclu definit de $(a, a + P)$, atunci graficul complet al acestei funcții se va obține prin repetări periodice ale graficului din ciclul respectiv. De aici tragem concluzia că dacă forma graficului unei funcții empirice se repetă după un anumit interval de timp, funcția respectivă ar putea să fie o funcție periodică.

În cadrul compunerii a două sau mai multe funcții periodice se constată că dacă P este o perioadă a funcției $f(x)$, atunci și numerele $2P$, $3P$, $4P$ etc vor fi de asemenea perioade, fapt care este confirmat și din definiția unei funcții periodice (vezi formula 2.14). De asemenea, și numerele $-2P$, $-3P$, $-4P$ etc vor fi perioade, acestea din urmă fiind importante pentru problemele care se referă la cercetarea anumitor evenimente din trecut.

Cea mai simplă funcție periodică are forma

$$y = A \cdot \sin(\omega \cdot x + \varphi) = A \cdot \sin\left(\frac{2\pi}{P} x + \varphi\right) \qquad (2.15)$$

unde A, ω și φ sînt mărimi constante. Această funcție se numește funcția armonică de amplitudine A, frecvență ω și fază inițială φ, originea acestor termeni fiind legată de oscilațiile armonice, adică de mișcările oscilatorii din mecanică.

Formula (2.15) se mai poate scrie

$$y = A(\sin \omega x \cdot \cos \varphi + \cos \omega x \cdot \sin \varphi)$$

în care notăm

$$a = A \cdot \sin \varphi, \qquad b = A \cdot \cos \varphi \qquad (2.16)$$

astfel că

$$y = a \cdot \cos \omega x + b \cdot \sin \omega x \qquad (2.17)$$

iar din (2.16) rezultă

$$A = \sqrt{a^2 + b^2}, \qquad \operatorname{tg} \varphi = \frac{a}{b} \qquad (2.18)$$

Din definiția perioadei unui fenomen periodic, avem

$$P = \frac{2\pi}{\omega} \qquad (2.19)$$

iar numărul $1/P$ este numărul de oscilații în unitatea de timp. Prin urmare, $\omega = 2\pi/P$ este frecvența, adică numărul de oscilații în intervalul de 2π unități.

Scanned with OKEN Scanner

În cercetările de acustică se întîlnește de asemenea noțiunea de armonică. O coardă fixată la amîndouă capetele poate vibra cu un anumit număr de frecvențe diferite. Cea mai joasă dintre frecvențe, ω_1, se numește frecvență fundamentală sau armonică fundamentală; ea are mărimea cea mai mică și corespunde perioadei celei mai mari (vezi formula 2.19) — perioada fundamentală. Celelalte frecvențe $2\omega_1$, $3\omega_1$, $4\omega_1$ etc se numesc armonice superioare.

Posibilitatea de a scrie o funcție sub forma (2.17) sau (2.15), adică posibilitatea de a reprezenta o curbă empirică periodică în funcție de sinusi și cosinusi — funcții armonice — ne este oferită de *analiza armonică*. Ea se aplică numai la fenomene cu cel puțin o perioadă bine determinată sau cel puțin o perioadă suspectată.

În anul 1807, matematicianul și inginerul francez Joseph Fourier (1768—1830) a arătat că orice funcție $f: R \to R$ care satisface anumite condiții se poate scrie sub forma unei serii cu un număr infinit de termeni sinusoidali, iar în 1829, Dirichlet (1805—1859) a arătat că o astfel de funcție trebuie să îndeplinească condițiile:

— f trebuie să fie periodică, adică pentru $x \in R$ să avem $f(x) = f(x + nP)$ unde P este perioada. Dacă f nu este periodică, dar este definită pe un domeniu finit, suma termenilor sinusoidali va converge spre f pe domeniul de definiție,

— f trebuie să fie uniformă,

— f trebuie ca cel puțin pe anumite domenii să fie continuă și să aibă cel mult un număr finit de discontinuități și salturi finite într-o perioadă,

— f trebuie să posede un număr finit de maxime și minime în decursul unei perioade.

— Integrala $\int_a^b f(x)\,dx$ trebuie să fie convergentă.

— Integrala $\int_a^b f(x)\,dx$ trebuie să fie finită cînd $f(x)$ este măsurată de la valoarea sa medie.

Condițiile lui Dirichlet sînt suficiente, dar nu toate dintre ele sînt și necesare, fapt care face ca teorema lui Fourier să fie valabilă pentru o clasă mare de funcții.

O funcție, f, care satisface condițiile enumerate mai sus, se poate scrie sub forma seriei Fourier

$$y = f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \qquad (2.20)$$

unde a_0, a_n și b_n sînt mărimi constante de la un ciclu la altul și se numesc *coeficienți Fourier*. Determinarea valorilor acestor constante

Scanned with OKEN Scanner

reprezintă așa-numita *analiză armonică, analiză Fourier,* sau *analiza formei undelor.*

În formula (2.20), primul termen, a_0, este diferit de zero cînd valoarea medie a funcției f, pe o perioadă,este diferită de zero. Acest termen se numește componenta continuă a funcției, iar funcția respectivă se numește funcție periodică cu termen constant. Dacă primul termen, a_0, este nul, avem de-a face cu o *funcție periodică alternativă oscilantă* sau *simetrică.*

Suma din membrul al doilea din (2.20) conține termeni de forma $a_n \cos nx$ și $b_n \sin nx$ care se numesc *armonice de ordinul n* în cosinus și respectiv sinus. Armonica de ordinul întîi se numește armonică fundamentală, iar amplitudinile armonicelor se determină prin mărimile a_n și b_n.

Există multe domenii de cercetare în care se utilizează analiza Fourier. Printre acestea amintim fenomenele meteorologice care se repetă cu o perioadă de un an sau mai mulți ani, mareele terestre și oceanice care au fost examinate cu scopul elaborării anumitor prognoze, periodicitatea cutremurelor de Pămînt (fără rezultat deosebit), variația anuală a microseismelor etc.

Aplicarea practică a unor funcții de forma (2.20) se simplifică foarte mult dacă se ține seama de faptul că sistemul trigonometric fundamental de funcții

$$1,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ \ldots,\ \cos nx,\ \sin nx, \tag{2.21}$$

care au aceeași perioadă, este alcătuit din „polinoame" ortogonale.

Ne propunem să verificăm condițiile de ortogonalitate a funcțiilor (2.21). În acest scop vom efectua următoarele calcule:

$$\int_{-\pi}^{\pi} \cos nx\, dx = \left[\frac{\sin nx}{n}\right]_{-\pi}^{\pi} = \frac{1}{n}\left[\sin(n\pi) - \sin(-n\pi)\right] = 0 \tag{2.22}$$

$$\int_{-\pi}^{\pi} \sin nx\, dx = \left[\frac{-\cos nx}{n}\right]_{-\pi}^{\pi} = \frac{1}{n}\left[-\cos(n\pi) + \cos(-n\pi)\right] = 0 \tag{2.23}$$

$$\int_{-\pi}^{\pi} \sin nx \cdot \sin nx\, dx = \int_{-\pi}^{\pi} \sin^2 nx\, dx = \int_{-\pi}^{\pi} \frac{1 - \cos 2nx}{2}\, dx =$$

$$= \frac{2\pi}{2} - \frac{1}{2}\int_{-\pi}^{\pi} \cos 2nx\, dx = \pi \tag{2.24}$$

Scanned with OKEN Scanner

$$\int_{-\pi}^{\pi} \cos\ nx \cdot \cos\ nx\ dx = \int_{-\pi}^{\pi} \cos^2 nx\ dx = \int_{-\pi}^{\pi} \frac{1+\cos 2nx}{2} dx =$$

$$= \frac{2\pi}{2} + \frac{1}{2}\int_{-\pi}^{\pi} \cos 2nx\ dx = \pi \qquad (2.24a)$$

Apoi ne folosim de formulele generale

$$\begin{cases} \cos\alpha \cos\beta = \frac{1}{2}[\cos(\alpha-\beta) + \cos(\alpha+\beta)] \\ \sin\alpha \sin\beta = \frac{1}{2}[\cos(\alpha-\beta) - \cos(\alpha+\beta)] \end{cases} \qquad (2.25)$$

astfel încît pentru orice n și m întregi, $n \neq m$, avem

$$\int_{-\pi}^{\pi} \cos nx \cdot \cos mx\, dx = \frac{1}{2}\int_{-\pi}^{\pi} [\cos(n+m)x + \cos(n-m)x]\, dx = 0 \qquad (2.26)$$

$$\int_{-\pi}^{\pi} \sin nx \cdot \sin nx\, dx = \frac{1}{2}\int_{-\pi}^{\pi} [\cos(n-m)x - \cos(n+m)x] dx = 0 \qquad (2.27)$$

În fine, utilizînd formula

$$\sin\alpha \cos\beta = \frac{1}{2}[\sin(\alpha+\beta) + \sin(\alpha-\beta)] \qquad (2.28)$$

pentru orice n și m întregi, avem

$$\int_{-\pi}^{\pi} \sin nx \cdot \cos\ mx\, dx = \frac{1}{2}\int_{-\pi}^{\pi} [\sin(n+m)x + \sin(n-m)x]\, dx = 0 \qquad (2.29)$$

Prin urmare, evaluarea integralelor de mai sus ne-a condus la determinarea proprietăților de ortogonalitate a sistemului de funcții definit de (2.21).

În activitatea practică, cum este aceea de prognoză, se întîmplă ca funcția $f(x)$, pe care trebuie s-o reprezentăm printr-o serie Fourier, să nu fie dată printr-o expresie analitică, ci prin tabele cu date nu-

Scanned with OKEN Scanner

merice sau grafic. În acest caz funcția respectivă o vom aproxima prin polinomul trigonometric care rezultă din (2.20), adică

$$f(x) = a_0 + a_1 \cos x + a_2 \cos 2x + \ldots + a_n \cos nx + \\ + b_1 \sin x + b_2 \sin 2x + \ldots + b_n \sin nx. \tag{2.30}$$

Pentru determinarea constantelor a_0, a_n și b_n vom înmulți în (2.30) succesiv cu fiecare din funcțiile sistemului (2.21) și apoi integrăm de-a lungul unei perioade 2π. În modul acesta, înmulțind în (2.30) cu „funcția" 1 și ținînd seamă de ortogonalitatea funcțiilor (2.21), avem

$$\int_0^{2\pi} f(x)\,dx = \int_0^{2\pi} a_0\,dx = 2\pi a_0$$

adică

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(x)\,dx \tag{2.31}$$

Cînd înmulțim în (2.30) cu funcția cos nx și integrăm, ținînd seamă de condițiile de ortogonalitate, rezultă

$$\int_0^{2\pi} f(x)\cos nx\,dx = \int_0^{2\pi} a_n \cos^2 nx\,dx = a_n\pi$$

ceea ce ne conduce la relația

$$a_n = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos nx\,dx \tag{2.32}$$

Procedînd la fel, prin înmulțirea cu sin nx și integrînd, din (2.30) rezultă

$$b_n = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin nx\,dx \tag{2.32a}$$

Prin urmare, formulele (2.31) — (2.32a) reprezintă expresiile generale ale coeficienților Fourier. Aceste formule însă nu pot fi aplicate în practică, deoarece nu cunoaștem expresia analitică a funcției $f(x)$ pentru ca să putem evalua integralele respective. De aceea va trebui să efectuăm un calcul aproximativ al constantelor Fourier atunci cînd dispunem de o serie de ordonate observate uniform de-a lungul unui ciclu. În acest scop va trebui să trecem de la formulele

Scanned with OKEN Scanner

exacte (2.31) — (2.32a) la niște formule aproximative, folosindu-ne de metodele integrării numerice.

Să considerăm segmentul $[0, 2\pi]$ împărțit în m părți egale prin punctele echidistante de pe axa absciselor

$$0, \frac{2\pi}{m}, 2\frac{2\pi}{m}, 3\frac{2\pi}{m}, \ldots, (m-1)\frac{2\pi}{m}, 2\pi$$

și să presupunem cunoscute valorile lui $f(x)$ pentru abscisele respective, adică să cunoaștem seria de ordonate

$$y_0, y_1, y_2, \ldots, y_{m-1}, y_m(=y_0).$$

Avînd în vedere faptul că funcția $f(x)$ este considerată periodică, putem scrie

$$\int_{x_0}^{x_m} f(x)\,dx = h\sum_{i=0}^{m-1} y_i, \tag{2.33}$$

(vezi de ex. Todoran, 1976, pg 201 și 203). În acest caz $h = 2\pi/m$ și

$$a_0 = \frac{1}{2\pi}\cdot\frac{2\pi}{m}\sum_{i=0}^{m-1} y_i.$$

Procedînd în mod analog și pentru celelalte constante, avem

$$\left.\begin{aligned} a_0 &= \frac{1}{m}\sum_{i=1}^{m-1} y_i \\ a_n &= \frac{2}{m}\sum_{i=0}^{m-1} y_i \cos nx_i, \quad \left(n \neq N\frac{m}{2}\right) \\ a_n &= \frac{1}{m}\sum_{i=0}^{m-1} y_i \cos nx_i, \quad \left(n = N\frac{m}{2}\right) \\ b_n &= \frac{2}{m}\sum_{i=0}^{m-1} y_i \sin nx_i, \end{aligned}\right\} \tag{2.34}$$

Pentru stabilirea celei de-a treia relații din (2.34), pornim de la faptul că, pentru $n = N\frac{m}{2}$, putem scrie

$$\sum_{i=0}^{m-1}\cos^2 nx_i = \sum_{i=0}^{m-1}\cos N\frac{m}{2}\cdot\frac{2\pi i}{m} = \sum_{i=0}^{m-1}\cos^2 N\pi i = m$$

Scanned with OKEN Scanner

unde avem

$$x_i = \frac{2\pi i}{m}.$$

Avînd în vedere produsul $Ni = N' =$ număr întreg, $\cos^2 N'\pi = 1$ pentru orice valoare a lui i. Prin urmare, pentru $n = N\frac{m}{2}$, avem

$$\sum_{i=0}^{m-1} \cos^2 nx_i = m.$$

Ne întoarcem acum la formula (2.30) pe care o înmulţim cu $\cos nx_i$, însumăm şi ţinem seamă de ortogonalitatea polinoamelor din (2.21) (vezi formulele 2.26 şi 2.29), putem scrie mai departe

$$\sum_{i=0}^{m-1} f(x_i) \cos nx_i = ma_n$$

de unde

$$a_n = \frac{1}{m} \sum_{i=0}^{m-1} y_i \cos nx_i, \qquad \left(n = N\frac{m}{2}\right)$$

cu $f(x_i) = y_i$.

Pentru cazul general, cînd perioada fundamentală a unei funcţii periodice este $P = 2\pi$, expresiile coeficienţilor Fourier şi a seriei corespunzătoare vor fi:

$$\left.\begin{aligned} a_0 &= \frac{1}{P}\int_{-P/2}^{P/2} f(t)\,dt \\ a_n &= \frac{2}{P}\int_{-P/2}^{P/2} f(t)\cos\frac{2n\pi t}{P}\,dt, \\ b_n &= \frac{2}{P}\int_{-P/2}^{P/2} f(t)\sin\frac{2n\pi t}{P}\,dt, \\ f(t) &= a_0 + \sum_{n=1}^{\infty}\left(a_n\cos\frac{2n\pi t}{P} + b_n\sin\frac{2n\pi t}{P}\right) \end{aligned}\right\} \qquad (2.34a)$$

unde drept variabilă am utilizat timpul t, care este de obicei variabilă independentă în aplicaţii.

Scanned with OKEN Scanner

Observații

— Coeficientul a_0 este definit de prima relație din (2.34) și, în condițiile amintite mai sus (intervalul $[0, 2\pi]$ împărțit într-un număr întreg de puncte), reprezintă media aritmetică a ordonatelor observate — a semnalelor înregistrate.

— În cazul cînd $f(x)$ este o funcție definită pe toată axa absciselor sau numai pe un anumit segment simetric față de origine, putem distinge două cazuri:

i) Dacă $f(x)$ este o funcție pară adică $f(-x) = f(+x)$, atunci pentru orice x, avem

$$\int_{-\pi}^{\pi} f(x)\mathrm{d}x = 2\int_{0}^{\pi} f(x)\mathrm{d}x.$$

ii) Dacă $f(x)$ este impară $f(-x) = -f(x)$, atunci

$$\int_{-\pi}^{\pi} f(x)\mathrm{d}x = 0$$

Din studiul unor astfel de funcții rezultă că produsul a două funcții pare sau a două funcții impare este o funcție pară, iar produsul dintre o funcție impară și o funcție pară este o funcție impară. În aceste condiții, dacă $f(x)$ este o funcție pară, știind că funcția $\cos nx$ este și ea pară din (2.32) vom avea

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\ \mathrm{d}x = \frac{1}{\pi}\int_{0}^{2\pi} f(x)\cos nx\ \mathrm{d}x = \frac{2}{\pi}\int_{0}^{\pi} f(x)\cos nx\ \mathrm{d}x,$$

pe cînd $\sin nx$ fiind o funcție impară, rezultă

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\ \mathrm{d}x = 0$$

Prin urmare, seria Fourier a unei funcții pare — simetrică față de origine — va conține numai termeni cu cosinuși, adică

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos nx.$$

În mod analog se arată că dacă funcția $f(x)$ este impară, atunci

$$f(x) = a_0 + \sum_{n=1}^{\infty} b_n \sin nx.$$

Scanned with OKEN Scanner

Prin urmare, dacă datele experimentale vor indica o funcție periodică pară sau impară, formula corespunzătoare se scrie cu mai multă ușurință, problema de prognoză putînd fi examinată cu un volum mai mic de calcule decît atunci cînd nu se știe nimic despre paritatea sau imparitatea funcției respective.

*

* *

Avînd în vedere faptul că în problemele de prognoză avem de-a face cu semnale înregistrate cu ajutorul numerelor reale, rezultă că funcția $f(x)$ este reală și, prin aceasta, toți coeficienții Fourier vor fi mărimi reale. Totuși, pentru o serie de considerații teoretice vom da aici și forma complexă a seriei Fourier. În acest scop, folosindu-ne de identitatea lui Euler, care leagă funcțiile trigonometrice de funcțiile exponențiale

$$e^{i\varphi} = \cos\varphi + i\sin\varphi$$

putem scrie imediat

$$\left.\begin{aligned} e^{i\varphi} &= \cos\varphi + i\sin\varphi \\ e^{-i\varphi} &= \cos\varphi - i\sin\varphi \end{aligned}\right\} \Rightarrow \cos\varphi = \frac{e^{i\varphi} + e^{-i\varphi}}{2},\ \sin\varphi = \frac{e^{i\varphi} - e^{-i\varphi}}{2i}$$

sau, în cazul general

$$\cos nx = \frac{e^{inx} + e^{-inx}}{2},\ \sin nx = i\,\frac{-e^{inx} + e^{-inx}}{2}, \qquad (2.35)$$

pe care le înlocuim în (2.20) și avem

$$f(x) = a_0 + \sum_{n=1}^{\infty}\left(a_n\,\frac{e^{inx} + e^{-inx}}{2} + ib_n\,\frac{-e^{inx} + e^{-inx}}{2}\right)$$

care se mai poate scrie și sub forma

$$f(x) = a_0 + \sum_{n=1}^{\infty}\left(\frac{a_n - ib_n}{2}\,e^{inx} + \frac{a_n + ib_n}{2}\,e^{-inx}\right). \qquad (2.36)$$

Dacă notăm

$$c_0 = a_0;\quad c_n = \frac{a_n - in_n}{2},\quad c_{-n} = \frac{a_n + ib_n}{2} \qquad (2.37)$$

$(n = 1, 2, 3, \ldots)$

atunci suma parțială de rang m a seriei (2.30) și deci a seriei (2.20) se poate scrie în modul următor

$$S_m(x) = c_0 + \sum_{n=1}^{m}\left(c_n e^{inx} + c_{-n}e^{-inx}\right) = \sum_{n=-m}^{m} c_n e^{inx}. \qquad (2.38)$$

Scanned with OKEN Scanner

Prin urmare, este firesc să folosim aproximația

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx} \tag{2.39}$$

care este *forma complexă a seriei Fourier* a lui $f(x)$.

Coeficienții c_n definiți prin formulele (2.37) se numesc coeficienți Fourier complecși ai funcției $f(x)$, iar forma lor generală este

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\, c^{-inx}\, dx \tag{2.40}$$

Formula (2.40) rezultă imediat după ce înlocuim pe (2.31) — (2.32a) în (2.37), unde se ține seamă și de (2.35).

Pentru o funcție reală (a_n și b_n reali) coeficienții c_n și c_{-n} sînt numere complexe conjugate, așa cum se poate vedea și din (2.37).

În cazul cînd funcția periodică $f(x)$ are perioada $2l$, mărimea x din formulele de mai sus se înlocuiește prin frecvența $\omega = \frac{\pi x}{l}$.

În aceste condiții formulele (2.39) și (2.40) devin

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \exp\left(\frac{in\pi x}{l}\right) \tag{2.41}$$

$$c_n = \frac{1}{2l} \int_{-l}^{l} f(x) \exp\left(-\frac{in\pi x}{l}\right) dx \tag{2.42}$$

unde am utilizat expresia *exp* prin care se vede că este vorba de e la o anumită putere, e fiind aici baza logaritmilor neperieni.

Să considerăm că funcția $f(x)$ este definită pentru orice x, iar în (2.42) să înlocuim variabila x printr-o variabilă fictivă u, în (2.41) avem

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{1}{2l} \int_{-l}^{l} f(u) \exp\left(-i\frac{\pi nu}{l}\right) \exp\left(i\frac{\pi nx}{l}\right) du$$

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{1}{2l} \int_{-l}^{l} f(u) \exp\left(in\pi\frac{x-u}{l}\right) du$$

Ne propunem acum să vedem ce devine suma din partea dreaptă a relației de mai sus. În acest scop vom nota

$$\lambda_1 = \frac{\pi}{l},\quad \lambda_2 = \frac{2\pi}{l},\quad \ldots,\quad \lambda_n = \frac{n\pi}{l};\quad \Delta\lambda_n = \lambda_{n+1} - \lambda_n = \frac{\pi}{l},$$

Scanned with OKEN Scanner

astfel că

$$f(x) = \lim \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \Delta\lambda_n f(u) \exp\left[i\lambda_n(x-u)\right] du$$

$$l \to \infty, \quad \Delta\lambda_n \to 0.$$

Dar dacă $\Delta\lambda_n \to 0$, putem scrie suma sub forma unei integrale

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\lambda \int_{-\infty}^{\infty} f(u) \exp\left[i\lambda\,(x-u)\right] du \tag{2.43}$$

care este *forma complexă a integralei Fourier.*

Integrala (2.43) se mai poate scrie

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(u)\, e^{-iu\lambda}\, du \cdot e^{ix\lambda}\, d\lambda$$

sau

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\lambda)\, e^{ix\lambda}\, d\lambda \tag{2.44}$$

unde am notat

$$F(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(u)\, e^{-iu\lambda}\, du \tag{2.45}$$

Formula (2.45) este *transformata directă Fourier,* iar (2.44) se numeşte *transformata inversă Fourier.* Aceste transformate ne permit, prin intermediul integrării, să efectuăm trecerea de la funcţia originală $f(u)$ care este dată, la imaginea ei $F(\lambda)$ şi apoi la $f(x)$ care este funcţia căutată. Funcţiile $F(\lambda)$ şi $f(x)$ formează un *cuplu* Fourier.

Pentru funcţii reale, formula (2.43) devine

$$f(x) = \frac{1}{\pi} \int_{0}^{\infty} d\lambda \int_{-\infty}^{\infty} f(u) \cos \lambda(x-u)\, du$$

sau, dezvoltînd cosinusul diferenţei respective, putem scrie mai departe

$$f(x) = \frac{1}{\pi} \int_{0}^{\infty} d\lambda \int_{-\infty}^{\infty} f(u)(\cos \lambda x \cdot \cos \lambda u + \sin \lambda x \cdot \sin \lambda u)\, du$$

care devine

$$f(x) = \frac{1}{\pi} \int_{0}^{\infty} \cos \lambda x \cdot d\lambda \int_{-\infty}^{\infty} f(u) \cos \lambda u\, du + \frac{1}{\pi} \int_{0}^{\infty} \sin \lambda x\, d\lambda \int_{-\infty}^{\infty} f(u) \sin \lambda u\, du$$

Scanned with OKEN Scanner

Vom defini aici *transformata cosinus*, $a(\lambda)$, prin relaţia

$$a(\lambda) = \int_{-\infty}^{\infty} f(u) \cos \lambda u \, du \tag{2.46}$$

iar *transformata sinus*, $b(\lambda)$, prin

$$b(\lambda) = \int_{-\infty}^{\infty} f(u) \sin \lambda u \, du \tag{2.47}$$

Funcţia $f(x)$ se poate scrie acum

$$f(x) = \frac{1}{\pi} \int_{0}^{\infty} \{a(\lambda) \cos \lambda x + b(\lambda) \sin \lambda x\} \, d\lambda \tag{2.48}$$

Funcţia de sub semnul integrală din (2.48) conţine suma a doi termeni cosinus şi sinus, astfel că ea poate fi adusă la forma

$$f(x) = \frac{1}{\pi} \int_{0}^{\infty} R(\lambda) \cos \{\lambda x + \varphi(\lambda)\} \, d\lambda \tag{2.49}$$

unde

$$R(\lambda) = \sqrt{\{a(\lambda)\}^2 + \{b(\lambda)\}^2}, \qquad \operatorname{tg} \varphi(\lambda) = -\frac{b(\lambda)}{a(\lambda)} \tag{2.50}$$

Observaţie. În unele lucrări λ este înlocuit prin ω (frecvenţa), iar abscisa x este înlocuită prin timpul t.

Toate consideraţiile pe care le-am prezentat mai sus au avut în vedere faptul că pe axa absciselor am considerat timpul, adică $f(x)$ era o funcţie de timp. Dar în examinarea a foarte multe probleme este mai comod dacă pe axa absciselor vom avea frecvenţa ω care se exprimă prin formula

$$\omega_k = \frac{2k\pi}{P} = k\omega_1, \qquad \omega_1 = \frac{2\pi}{P} \tag{2.51}$$

$(k = 1, 2, 3, \ldots)$

Aici k arată numărul de ordine al unei armonice, ω_1 fiind *frecvenţa fundamentală*. Prin urmare, pentru prima armonică avem $\omega_1 = 2\pi/P$, pentru armonica a doua avem $\omega_2 = 2.2\pi/P$ etc, astfel că pe axa absciselor în locul timpului continuu putem lua valorile discrete ale frecvenţelor.

Utilizarea frecvenţelor în locul timpului pentru variabila independentă este avantajoasă atît pentru faptul că este mai comodă

Scanned with OKEN Scanner

pentru anumite operații matematice, cît și pentru aceea că o reprezentare de acest fel poate fi mai direct legată de fenomenele cercetate, multe fenomene fiind înregistrate în domeniul frecvențelor.

Dacă ne referim acum la un semnal periodic care se repetă după o perioadă P și care satisface condițiile Dirichlet, seria Fourier poate fi scrisă și sub forma

$$x(t) = f(t) = c_0 + \sum_{k=1}^{\infty} c_k \sin(k\omega t + \varphi_k) \tag{2.52}$$

unde

$$c_0 = \frac{1}{P}\int_{-P/2}^{P/2} x(t)\mathrm{d}t, \quad c_k^2 = a_k^2 + b_k^2 \tag{2.53}$$

$(k = 1, 2, 3, \ldots)$

$$a_k = \frac{2}{P}\int_{-P/2}^{P/2} x(t)\cos k\omega t\, \mathrm{d}t, \qquad b_k = \frac{2}{P}\int_{-P/2}^{P/2} x(t)\sin k\omega t\, \mathrm{d}t. \tag{2.54}$$

Aici mărimile c_k formează *spectrul amplitudinilor*, iar mulțimea

$$\varphi_k = \operatorname{arc\,tg}\frac{a_k}{b_k} \tag{2.55}$$

constituie *spectrul fazelor*. De obicei, în practică se utilizează numai spectrul amplitudinilor, care dealtfel se numește simplu *spectru*. Prin urmare, spectrul unui semnal periodic va fi reprezentat prin seria de numere c_k. Reprezentarea grafică va scoate în evidență prezența anumitor linii verticale egal distanțate, a căror lungime este egală cu mărimile c_k, care vor fi situate la abscisele ω_k (vezi figura 2.1). Un astfel de spectru se numește spectru *discret*.

Funcția neperiodică $x(t)$ poate fi considerată ca un caz particular al funcțiilor periodice cu perioada P care tinde spre infinit.

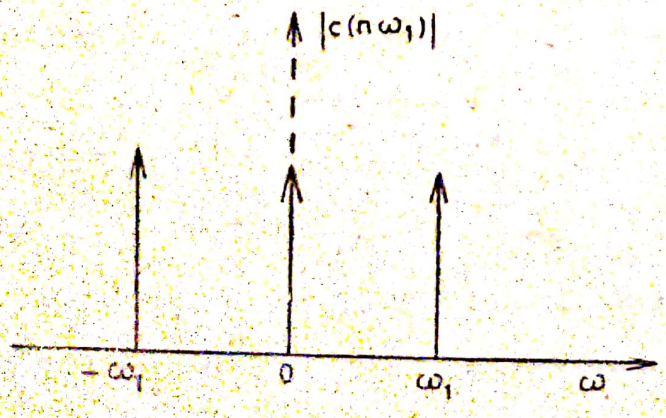


Fig. 2.1. Reprezentarea spectrului amplitudinilor

Pe baza formulei $\omega_1 = 2\pi/P$ care definește frecvența fundamentală, cînd $P \to \infty$, $\omega_1 \to 0$, ceea ce înseamnă că distanța dintre liniile spectrale tinde spre zero, iar spectrul de linii tinde spre un spectru continuu.

Scanned with OKEN Scanner

În domeniul frecvențelor, seria Fourier din (2.41) devine

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega_1 x} \tag{2.56}$$

unde

$$c_n = c(n\omega_1) = \frac{1}{P} \int_{-P/2}^{P/2} f(x)\ e^{-in\omega_1 x}\, dx \tag{2.57}$$

Aici $c_n = c(n\omega_1)$ definește transformata Fourier a semnalului $f(x)$ și este o reprezentare în domeniul frecvențelor.

Relația (2.56) reprezintă sinteza semnalului $f(x)$, iar relația (2.57) este analiza Fourier a lui $f(x)$.

Dacă o funcție are aceeași formă cînd este reprezentată în domeniul timpului ca și atunci cînd este reprezentată în domeniul frecvențelor, ea se numește *funcție autoreciprocă*.

Să considerăm seria Fourier definită în domeniul frecvențelor prin intermediul formulelor (2.56) și (2.57) și să admitem că funcțiile aperiodice provin din funcțiile periodice atunci cînd perioada P crește nelimitat. În acest caz, din formula (2.57) care [vezi (2.53)] definește spectrul, putem scrie mai departe

$$c(n\omega_1) = \frac{1}{P} \int_{-P/2}^{P/2} f(t)e^{-in\omega_1 t}\, dt = \frac{1}{2\pi}\,\omega_1 \int_{-P/2}^{P/2} f(t)e^{-in\omega_1 t}\, dt \tag{2.58}$$

Dacă P tinde către infinit, ω_1 devine o mărime diferențială $d\omega$, iar variabila discretă $n\omega_1$ tinde spre variabila continuă ω.

Prin urmare,

$$2\pi \frac{c(n\omega_1)}{\omega_1} = \int_{-P/2}^{P/2} f(t)e^{-in\omega_1 t}\, dt$$

va tinde spre o limită $S(\omega)$, exprimată prin relația

$$S(\omega) = \lim_{P\to\infty} 2\pi \frac{c(n\omega_1)}{\omega_1} = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\, dt \tag{2.59}$$

Funcția $S(\omega)$ astfel definită se numește *densitate spectrală* și reprezintă limita raportului lui $c(n\omega_1)$ împărțit la frecvența ω_1 cînd $\omega_1 = d\omega$.

Am prezentat aici doar cîteva dintre problemele legate de teoria semnalelor, cîteva probleme necesare la elaborarea unei prognoze asupra căreia vom reveni în capitolele următoare. Pentru o examinare mai detaliată a unor astfel de probleme recomandăm lucrările următorilor autori: Spătaru (1965), Gîrlașu (1978) și Stanomir și Stănășilă (1980).

Scanned with OKEN Scanner

CAPITOLUL III

NOȚIUNI DE TEORIA CORELAȚIEI

În primul capitol am scos în evidență necesitatea cunoașterii corelației dintre diferiți parametri prin care se caracterizează modul de manifestare a fenomenelor naturale și necesitatea elaborării unei prognoze științifice. În acest scop ne-am referit la unele dintre cele mai capricioase manifestări ale naturii : cutremure de Pământ, erupții vulcanice, activitatea meteorologică și la unele probleme din domeniul vieții sociale : stabilirea prețurilor, elaborarea planurilor economice, reducerea mortalității etc. Aceste fenomene trebuie să fie urmărite cu multă atenție pentru a se putea pune în evidență unele corelații care să fie utilizate în probleme de prognoză.

De asemenea, am arătat că problemele de corelație și prognoză trebuie să le cercetăm luînd în considerare legătura dintre diferite fenomene precum și interdependența dintre diferiți factori care definesc unul și același fenomen. În modul acesta devine posibilă stabilirea unui tablou al condițiilor în care se desfășoară un anumit fenomen natural sau proces social, fapt care ar putea duce la modelarea matematică corespunzătoare. Astfel de probleme se cercetează cu mult succes prin metodele statisticii matematice, unde capitolul privind teoria corelației are o pondere deosebit de importantă. Asupra unor astfel de probleme vom insista în cele ce urmează.

Fenomenele naturale sau procesele sociale se găsesc într-o interdependență reciprocă, unele dintre acestea putînd fi cunoscute, iar altele nu. Tocmai de aceea, în cercetarea legăturii reciproce dintre două fenomene apar dificultăți mari deoarece aici pot să existe și cauze (fenomene) necunoscute, relațiile cauzale fiind mascate de acțiuni nebănuite. În astfel de situații este util a determina gradul de corelare și apoi de a examina „separat" unele dintre cauzele respective. Cu alte cuvinte, trebuie să ne referim și la corelația posibilă dintre diferiți factori, evenimente, atribute sau caracteristici care ar putea avea o influență, cel puțin parțială, asupra datelor experimentale. În astfel de cazuri rezultatele obținute vor constitui o suprapunere a mai multor cauze, unele dintre acestea putînd fi doar bănuite.

Statistica matematică este o ramură a științelor matematice și se ocupă și de rezolvarea unor probleme pe care le ridică activitatea practică. În mod deosebit ea se ocupă de prelucrarea și interpretarea datelor empirice obținute prin observarea fenomenelor de masă și

Scanned with OKEN Scanner

încearcă să stabilească anumite legități statistice care se manifestă în fenomenele naturale cercetate sau a proceselor sociale. Aceste chestiuni fiind direct legate de scopul expunerii noastre — elaborarea unei prognoze eficiente — considerăm util să prezentăm aici, foarte pe scurt, unele probleme de statistică matematică care, într-un mod sau altul, ne pot duce la o prognoză îmbunătățită. Cu alte cuvinte, ne vom ocupa de teoria corelației, cu ajutorul căreia se pot pune în evidență trăsături comune sau interdependența dintre fenomenele naturale și procesele sociale.

3.1. Colectivitate statistică

În cadrul problemelor de cercetare științifică, de foarte multe ori statistica matematică utilizează noțiunea de colectivitate statistică care, din punct de vedere matematic, se definește ca *o mulțime de obiecte fizice sau abstracte, numite unități ale colectivității, care au un număr de însușiri esențiale comune.*

Omogenitatea unităților unei colectivități este caracterizată prin faptul că între unitățile componente există o legătură comună. Această legătură este specifică unui anumit grup considerat și se manifestă printr-o mărime stabilă cum ar fi, de exemplu, valoarea medie, determinînd în acest mod conținutul legității principale. O astfel de stabilitate se numește *legitatea statistică* și poate deveni cognoscibilă atunci cînd numărul cazurilor observate devine din ce în ce mai mare.

Proprietatea urmărită în cercetarea unei colectivități se numește *caracteristică.* Aceasta se poate prezenta sub aspect cantitativ care se exprimă numeric sau sub aspect calitativ care se exprimă, de obicei, prin *da* sau *nu.* Aspectul calitativ al unei caracteristici se poate evalua și prin codificare sau stabilirea unor trepte arbitrare.

Exemple de colectivități statistice se pot da foarte multe și din cele mai diferite domenii, începînd cu macrocosmosul, societatea umană și terminînd cu microcosmosul.

— Stelele de un anumit tip spectral. Caracteristici comune: temperatura, masa, luminozitate etc.

— Populația unui oraș sau a unei țări la care putem evidenția următoarele caracteristici: vîrsta, profesiunea, înălțimea, sexul etc.

— Piesele produse de o mașină într-un interval de timp dat. Caracteristici comune: dimensiunile, rezistența la anumite probe, prețul de cost etc.

În cadrul aceleiași colectivități, o caracteristică se poate schimba de la o unitate la alta. De aceea, la fiecare caracteristică teoria probabilităților asociază o variabilă aleatoare care poate să fie discretă sau continuă. Este vorba de o variabilă aleatoare discretă atunci cînd

Scanned with OKEN Scanner

ea poate lua numai anumite valori cum este, de exemplu, numărul de copii dintr-o familie. O familie poate avea un număr întreg de copii: 0, 1, 2, ...

O variabilă continuă poate lua orice mărime dintr-un interval dat. Astfel, de exemplu, producția medie a unui teren arabil sau cantitatea de apă căzută într-o zi sînt variabile continue.

O colectivitate statistică, prin caracteristicile sale, în principiu, se supune anumitei legități specifice care există în mod obiectiv. Trăsăturile caracteristice ale colectivității decurg din legitatea respectivă, sînt legate de ea și nu pot fi înțelese dacă nu se ține seamă de această legătură.

Cercetarea unei caracteristici statistice (variabilă aleatoare) se face, în general, prin metoda „sondajului" care, uneori, este singura metodă posibilă și economică. Este vorba de cercetarea unei submulțimi, a unor eșantioane, părți reduse din colectivitatea generală. Teoria statistică arată în ce condiții această metodă este valabilă și eficientă, mai ales cînd volumul de date utilizate este suficient de mare.

După cum am mai amintit, scopul expunerii noastre este acela de a cerceta condițiile în care am putea ajunge la o prognoză cît de cît demnă de încredere. În acest scop trebuie să avem în vedere faptul că domeniile principale care ne preocupă vor fi acelea ale fenomenelor naturale și cel al proceselor sociale, iar cercetările respective trebuie să țină seama de caracterul evolutiv al acestora. În plus, atît fenomenele naturale cît și procesele sociale trebuie privite în complexitatea lor și în interdependența dintre ele. Fiecare fenomen natural se manifestă în mijlocul altor fenomene naturale, un proces social este în mijlocul altor procese sociale și, în același timp, este supus și fenomenelor naturale. De aceea, modelarea matematică a acestora este destul de complicată și se cer date experimentale foarte numeroase.

Atît fenomenele naturale cît și procesele sociale, pe lîngă complexa întrepătrundere, mai sînt supuse unei evoluții care la prima vedere poate fi considerată probabilistă (aleatoare). De aceea, pentru cunoașterea modului de evoluție probabilă în viitor — prognoza unui anumit fenomen — trebuie să ne bazăm pe cunoașterea evoluției trecute și prezente, iar experiența unui mare număr de cercetări a dus la concluzia că între diferite mărimi variabile, între anumite caracteristici ale membrilor unei colectivități statistice sau ale membrilor diferitelor colectivități statistice, pot exista următoarele tipuri de relații:

a) *Relația de dependență*, Y depinde de X sau invers. Aici o modificare a unei variabile duce la o modificare a celei de-a doua variabile. În cazul unei astfel de relații s-ar putea aminti existența relației cauzale în care o variabilă este *cauza*, iar cealaltă este *efectul*, cauza fiind un fenomen sau un complex de fenomene care provoacă, generează sau determină un alt fenomen — *efectul*. Prin urmare cauza precede

Scanned with OKEN Scanner

iar efectul succede, efectul fiind produsul cauzei. Operația logică prin care efectul este dedus din cauză se numește *inferență cauzală.*

b) *Relația de interdependență*, Y depinde de X și X depinde de Y. În acest caz modificarea unei variabile provoacă modificare celei de a doua variabile, iar modificarea acesteia din urmă are influență asupra primei variabile.

c) *Relația de tranziție*, X se transformă — parțial sau total în Y și invers.

d) *Corelația statistică* sau covarianța pentru X există întotdeauna Y și invers, este o relație reciprocă dintre două variabile, dintre care una în mod logic apelează la alta și, pe baza analizei datelor experimentale, se poate semnala o asociere între ele.

e) *Relația stochastică*, dacă X atunci foarte probabil și Y sau invers.

f) *Relația de transmitere a informațiilor*, Y recepționează informații de la X sau invers.

g) *Relația cronologică*, dacă X atunci mai tîrziu Y sau invers.

Avînd în vedere faptul că elementele de prognoză, la care ne vom referi în continuare, pot fi utilizate și în cadrul științelor sociale, mai amintim aici cîteva noțiuni care pot fi legate de o colectivitate statistică.

Un *sistem social* reprezintă unitatea dialectică dintre *structura socială* și *procesul social*, unde

Structura socială este un sistem de colectivități umane care au caracteristici diferite, dar sînt astfel reunite încît ele constituie un organism social complex, iar

Procesul social este desfășurarea în timp a unui complex de fapte sociale.

3.2. Frecvența unui eveniment dat

Să considerăm o colectivitate statistică care conține n unități elementare pentru care, în mod empiric, s-a înregistrat numărul k de unități ce posedă o anumită caracteristică. De exemplu, dacă n reprezintă numărul de familii dintr-un oraș iar n_k este numărul de familii care au copii, atunci n_k se numește frecvență absolută iar

$$f_{k,n} = \frac{n_k}{n} \tag{3.0}$$

se numește frecvență relativă a familiilor care au copii în orașul respectiv.

O primă sistematizare a datelor asociate unei caracteristici se face prin scrierea lor în formă de tabel simplu, aici datele fiind trecute în ordinea în care ele au fost obținute. Plecînd de la acest tabel simplu,

Scanned with OKEN Scanner

datele empirice pot fi grupate în clase „omogene", după anumite criterii, ajungîndu-se apoi la un tabel sistematizat.

Să considerăm o colectivitate la care ne interesează o anumită caracteristică. În acest caz cele n rezultate obținute experimental le grupăm în N clase, numărul de elemente care intră într-o clasă reprezentînd frecvența absolută. Într-o astfel de situație tabelul sistematizat va fi de forma

Clasa	c_1	c_2	...	c_i	...	c_N
frecvența	n_1	n_2	...	n_i	...	n_N
	a_1	a_2	...	a_i	...	a_N

Aici n_i este frecvența absolută a clasei de indice i, iar a_i este valoarea corespunzătoare mijlocului intervalului de valori din clasa respectivă.

În practică, de obicei, în afara frecvenței absolute se folosește *frecvența relativă* care, cu notațiile de mai sus, se definește prin relația

$$f_i = \frac{n_i}{n}, \qquad i = \overline{1, N} \tag{3.1}$$

unde $n = \sum_{i=1}^{N} n_i$.

Dacă X este o caracteristică statistică iar $x_1, x_2, \ldots, x_n$ sînt datele empirice, atunci

$$X \begin{pmatrix} a_i \\ \frac{n_i}{n} \end{pmatrix}$$

reprezintă o *distribuție statistică* sau *empirică*.

Pe de altă parte, dacă dintr-o serie de înregistrări sau măsurări, efectuate asupra unui fenomen dat într-un complex de condiții, dispunem de un număr destul de mare de rezultate care oscilează tot timpul în jurul unei valori medii constante, spunem că fenomenul respectiv are o probabilitate de realizare.

Raportul dintre numărul n_k de realizări ale fenomenului (cazuri reușite) și numărul total n de probe (totalitatea cazurilor posibile) definește probabilitatea empirică

$$p = \frac{n_k}{n}. \tag{3.2}$$

Scanned with OKEN Scanner

În general frecvențele relative nu coincid cu probabilitățile teoretice, dar le aproximează pe acestea în cazul cînd s-a efectuat un număr mare de probe. Aceste frecvențe joacă un rol deosebit de important în aplicațiile practice.

3.3. Valoarea medie și dispersia unei distribuții empirice

Dacă într-o serie de n determinări, o caracteristică (variabilă aleatoare) X a apărut într-o ordine oarecare cu valorile

$$\underbrace{x_1,\ x_1,\ \ldots,\ x_1}_{n_1},\quad \underbrace{x_2,\ x_2,\ \ldots,\ x_2}_{n_2}\quad \ldots\quad \underbrace{x_k,\ x_k,\ \ldots,\ x_k}_{n_k}$$

atunci, pentru *valoarea medie*, $\bar{x}$, putem scrie

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \ldots + n_kx_k}{n} \tag{3.3}$$

unde $n = n_1 + n_2 + \ldots + n_k$.

Utilizînd noțiunea de frecvență relativă, adică folosind relația (3.2), din (3.3) avem mai departe

$$\bar{x} = \sum_{i=1}^{n} x_i f_i; \qquad f_i = \frac{n_i}{n}, \qquad (i = \overline{1, n}) \tag{3.4}$$

În mod analog, pentru dispersie putem scrie

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 n_i}{n} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i^2 - \bar{x}^2 \tag{3.5}$$

unde

$$\sigma = \sqrt{\sigma^2} \tag{3.6}$$

este abaterea medie pătratică numită și abatere sau deviație standard.

Formula (3.5) se mai poate scrie și sub forma

$$\sigma^2 = \sum_{i=1}^{n} (x_i^2 - 2x_i\bar{x} + \bar{x}^2)\frac{n_i}{n} \qquad (i = \overline{1, k}) \tag{3.7}$$

Este ușor de arătat că pentru $n_i = 1$ formulele (3.3) și (3.5) devin

$$m_0 = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.8}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 \tag{3.9}$$

Scanned with OKEN Scanner

3.4. Corelația

Pînă acum ne-am referit la determinarea unei caracteristici a unei colectivități statistice și a indicatorului numeric asociat acestora; este vorba de valoarea medie $\bar{x} = m_0$ a unei serii de x_i determinări. În cele ce urmează vom căuta să extindem problema noastră și să examinăm două sau mai multe caracteristici, fapt care ne apropie și mai mult de problemele puse prin activitatea practică. Considerăm aici cazul în care colectivitatea statistică va fi caracterizată prin două caracteristici care se pot prezenta sub următoarele aspecte:

— cele două caracteristici sînt independente între ele,

— între cele două caracteristici există o corelație, o dependență probabilistică, care încă nu constituie o relație funcțională propriu-zis, și

— între cele două caracteristici poate exista o corelație atît de puternică încît aceasta se transformă într-o relație funcțională. În acest din urmă caz este vorba de o dependență deterministă.

În problemele cercetate de statistică există multe cazuri de corelație și pot fi date multe exemple, noi amintim aici doar următoarele corelații:

— între notele studenților de la facultatea de matematică și cea de fizică,

— între temperaturile zilnice de la amiază la Mamaia și Întorsura Buzăului,

— între înălțimile taților și cele ale fiilor lor,

— între cantitatea de apă provenită din precipitațiile atmosferice căzute într-o lună și temperatura medie respectivă,

— între magnitudinile aparente și mișcările proprii ale stelelor etc.

Să considerăm o colectivitate statistică caracterizată prin caracteristicile X și Y. Efectuînd o serie de determinări experimentale asupra acestei colectivități, putem întocmi tabela datelor respective

$$\begin{array}{c|l} X & x_1,\ x_2,\ x_3,\ \ldots,\ x_n \\ Y & y_1,\ y_2,\ y_3,\ \ldots,\ y_n \end{array} \qquad (A)$$

Repartiția empirică a celor două variabile X și Y se obține pe cale grafică în sistemul de axe xOy, unde vom reprezenta punctele de coordonate x_i și y_i. Un ansamblu de astfel de puncte se numește *cîmp de corelație, tabel de corelație sau nor statistic*.

Dacă punctele $M_i(x_i, y_i)$ sînt distribuite de-a lungul unei fîșii care, în general, urmează o curbă determinată, spunem că între mărimile respective există o dependență funcțională (vezi figura 3.1 a).

Scanned with OKEN Scanner

Dacă punctele $M_i(x_i, y_i)$ nu arată o dependență funcțională strictă, dar există o tendință ca valorile lui Y să depindă de cele ale lui X deși nu în mod riguros, între mărimile X și Y există o corelație (vezi fig. 3.1 b).

În cazul cînd între X și Y nu există nici un fel de dependență, cîmpul de distribuție se va prezenta asemănător cu acela din figura 3.1 c. Cele două caracteristici X și Y sînt independente.

Într-un caz particular, dependența corelațională se poate transforma în dependență funcțională cînd, așa cum s-a amintit mai sus, punctele cîmpului de distribuție se situează pe o anumită curbă, eventual pe o linie dreaptă. Dar, în general, dependența corelațională este totdeauna mai mult sau mai puțin abătută de la dependența funcțională. De aici apare problema de a stabili în mod numeric în ce măsură dependența corelațională se apropie sau se îndepărtează de dependența funcțională.

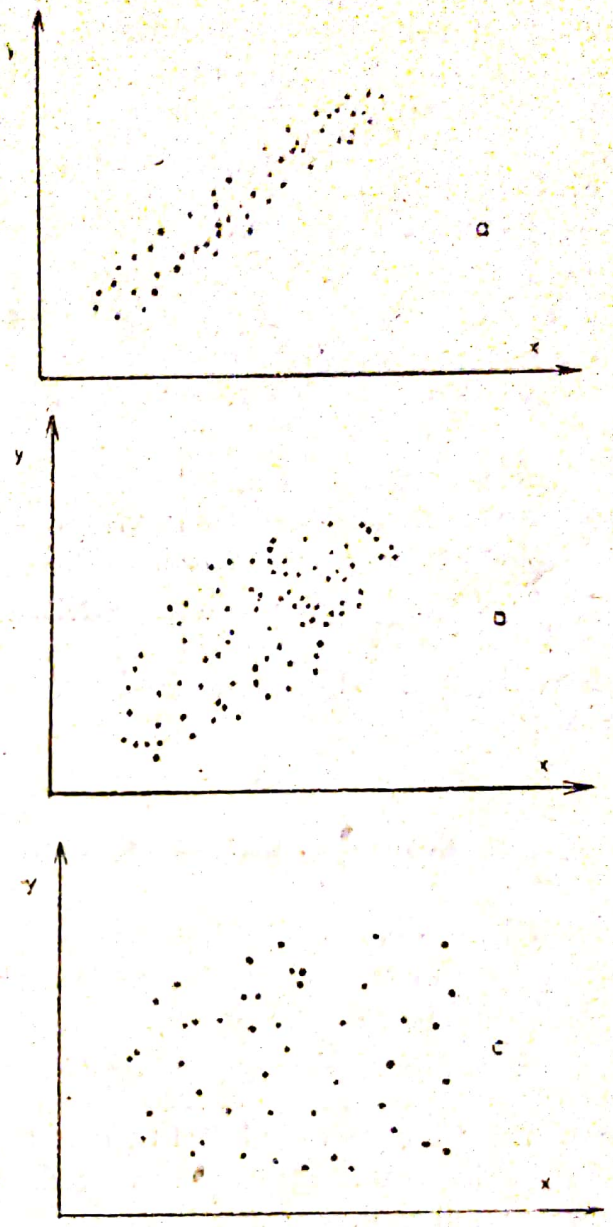


Fig. 3.1. Trei aspecte ale cîmpului de distribuție: a) Dependență funcțională b) Dependență corelațională și c) Lipsa oricărui fel de dependență.

În foarte multe cazuri, din observarea fenomenelor naturale sau a proceselor sociale, fără a cunoaște natura exactă a acestora și nici cauzele prin care este pusă în evidență o anumită caracteristică, se pot trage concluzii foarte importante prin examinarea corelației dintre aceste trăsături și alte evenimente cum ar fi de exemplu scurgerea timpului pentru unele variații periodice. În astfel de cazuri, analiza corelațională poate aduce contribuții valoroase deoarece aceasta este o metodă de cercetare a fenomenelor naturale și sociale și se bazează pe descompunerea unui întreg în elementele lui componente și pe analiza relațiilor statistice dintre acestea. În acest mod se poate aprecia existența unei relații statistice între două

Scanned with OKEN Scanner

sau mai multe variabile, adică, în astfel de cazuri se vorbește despre corelații dintre mărimile care indică o dependență reciprocă. Este vorba deci de o prezentare a unor legături stochastice sau corelate în locul unor dependențe funcționale.

În cazul cînd se consideră numai două variabile, se pot stabili corelații simple, iar cînd numărul variabilelor este mai mare decît doi, avem *corelații multiple*.

Din punct de vedere al formei unei relații statistice, spunem că ea este *directă* atunci cînd creșterea unei variabile, x, duce la o creștere a celeilalte variabile, y. Dar dacă la o creștere a lui x îi corespunde o descreștere a lui y, atunci avem o corelație *inversă*.

Relațiile statistice definite prin corelație ar putea avea aspect *liniar* sau *neliniar*, prin urmare se vor putea distinge *corelații liniare* și *neliniare* sau *curbilinii*.

În fine, în cazul cînd se utilizează mai multe variabile vom putea distinge *corelație parțială* cînd se consideră constante unele variabile și *corelație totală* cînd se iau în considerare variațiile tuturor mărimilor variabile.

3.5. Corelația liniară. Ecuația dreptei de regresie

Cînd între variabilele X și Y există o corelație și nu o dependență funcțională, atunci nu va fi vorba de a găsi „cea mai bună" valoare a lui Y care să corespundă unei valori a lui X, în acest caz ne vom referi numai la *cea mai probabilă valoare* a lui Y în jurul căreia ar putea fi distribuite valorile observate, în conformitate cu o anumită lege de distribuție a frecvențelor. Desigur, cu cît valorile observate vor fi mai apropiate de valoarea cea mai probabilă, cu atît legătura dintre X și Y este mai bine pusă în evidență. Această considerație se află la baza diferitelor mărimi numerice prin care se determină gradul de corelație.

Dintre tipurile de corelație amintite mai sus, vom examina aici tipul de corelație liniară. În acest scop vom considera cazul cînd o astfel de dependență este „indicată" de reprezentarea grafică a punctelor $M_i(x_i, y_i)$. Prin urmare avem motive să scriem

$$a + b \cdot x_k = y_k \tag{3.10}$$

Aici a și b sînt două mărimi constante care se determină din condiția ca mărimile y_k care se calculează cu ajutorul lor să aibă proprietatea ca suma pătratelor diferențelor $(y_k)_{obs} - (y_k)_{calc}$ să fie minimă. Este vorba de aplicarea metodei celor mai mici pătrate la

Scanned with OKEN Scanner

rezolvarea unui sistem de n ecuații $(k = \overline{1, n})$ cu două necunoscute. În aceste condiții (3.10) se poate scrie

$$S = \sum_{k=1}^{n} (a + b \cdot x_k - y_k)^2 \tag{3.11}$$

iar valoarea minimă a acesteia are loc atunci cînd

$$\frac{\partial S}{\partial a} = 2 \sum_{k=1}^{n} (a + b \cdot x_k - y_k) = 0 \tag{3.12}$$

$$\frac{\partial S}{\partial b} = 2 \sum_{k=1}^{n} (a + b\, x_k - y_k) x_k = 0$$

Aici înmulțim cu $f_k = \frac{n_k}{n}$ și ținem seama de definiția valorii medii (3.4) și de aceea a dispersiei (3.5), astfel că putem scrie mai departe

$$a + b \cdot \bar{x} - \bar{y} = 0 \tag{3.13}$$

$$a \cdot \bar{x} + b(\sigma_x^2 + \bar{x}^2) - \nu_{11} = 0 \tag{3.14}$$

unde am notat

$$\nu_{11} = \frac{1}{n} \sum_{k=1}^{n} x_k y_k n_k \tag{3.15}$$

iar σ_x^2 este dispersia mărimilor x.

În ecuația (3.14) adunăm și scădem produsul $\bar{x} \cdot \bar{y}$, astfel încît putem scrie mai departe

$$a \cdot \bar{x} + b \cdot \bar{x}^2 - \bar{x} \cdot \bar{y} + b\sigma_x^2 - \nu_{11} + \bar{x} \cdot \bar{y} = 0 \tag{3.16}$$

Suma primilor trei termeni în (3.16) este nulă (vezi 3.13) astfel că avem

$$b \cdot \sigma_x^2 = \mu_{11} \tag{3.17}$$

unde am notat

$$\mu_{11} = \nu_{11} - \bar{x} \cdot \bar{y} \tag{3.18}$$

În modul prezentat aici am obținut un sistem de două ecuații cu cele două necunoscute a și b. Prin urmare, ținînd seama de (3.13) și (3.17); avem

$$\left.\begin{aligned} a + b \cdot \bar{x} &= \bar{y} \\ b \cdot \sigma_x^2 &= \mu_{11} \end{aligned}\right\} \tag{3.19}$$

Scanned with OKEN Scanner

sau

$$\left.\begin{aligned} b &= \frac{\mu_{11}}{\sigma_x^2} \\ a &= \bar{y} - \frac{\mu_{11}}{\sigma_x^2}\bar{x} \end{aligned}\right\} \tag{3.20}$$

Revenind la forma generală care leagă cele două mărimi X și Y, din (3.10) avem

$$Y = a + b \cdot X \tag{3.21}$$

sau, înlocuind pe a și b cu expresiile lor din (3.20), avem

$$Y - \bar{y} = \frac{\mu_{11}}{\sigma_x^2}(X - \bar{x}) \tag{3.22}$$

unde am utilizat următoarele notații

$$\left.\begin{aligned} \bar{x} &= \frac{1}{n}\sum_{k=1}^{n} n_k x_k\,; & \bar{y} &= \frac{1}{n}\sum_{k=1}^{n} n_k y_k \\ \sigma_x^2 &= \frac{1}{n}\sum_{k=1}^{n} n_k x_k^2 - \bar{x}^2\,; & \mu_{11} &= \frac{1}{n}\sum_{k=1}^{n} n_k x_k y_k - \bar{x}\cdot\bar{y} \end{aligned}\right\} \tag{3.23}$$

În cazul cînd punctele $M_k(x_k, y_k)$ au toate aceeași probabilitate $p = 1$, avem

$$\left.\begin{aligned} \bar{x} &= \frac{1}{n}\sum_{k=1}^{n} x_k & \bar{y} &= \frac{1}{n}\sum_{k=1}^{n} y_k \\ \sigma_x^2 &= \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \bar{x}^2 & \mu_{11} &= \frac{1}{n}\sum_{k=1}^{n} x_k y_k - \bar{x}\cdot\bar{y} \end{aligned}\right\} \tag{3.24}$$

Procedînd exact în același mod pentru X ca funcție de Y, avem

$$X - \bar{x} = \frac{\mu_{11}}{\sigma_y^2}(Y - \bar{y}) \tag{3.25}$$

unde

$$\sigma_y^2 = \frac{1}{n}\sum_{k=1}^{n} y_k^2 - \bar{y}^2 \tag{3.26}$$

pentru $n_k = 1$, și

$$\sigma_y^2 = \frac{1}{n}\sum_{k=1}^{n} n_k y_k^2 - \bar{y}^2 \tag{3.27}$$

Scanned with OKEN Scanner

pentru cazul $n_k \neq 1$ cînd mărimile x_k și y_k sînt determinate cu ponderi diferite.

În cazul cînd valorile x_i sau y_i ar avea un caracter ciclic, datele respective ar trebui incluse în tabele corespunzătoare.

Dăm aici, schematic, tabela X cînd la o valoare a lui x corespunde un grup de valori y, iar în tabela Y reproducem grupele de valori ale lui x pentru un y oarecare.

$$\begin{array}{c|cccc|c} x^{(1)} & y_{11}, & y_{12}, & \dots, & y_{1m} & \bar{y}_{x1} \\ x^{(2)} & y_{21}, & y_{22}, & \dots, & y_{2m} & \bar{y}_{x2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x^{(l)} & y_{l1}, & y_{l2}, & \dots, & y_{lm} & \bar{y}_{xl} \end{array} \qquad (X)$$

În mod cu totul analog se construiește tabela (Y).

$$\begin{array}{c|cccc|c} y^{(1)} & x_{11}, & x_{12}, & \dots, & x_{1l} & \bar{x}_{y1} \\ y^{(2)} & x_{21}, & x_{22}, & \dots, & x_{2l} & \bar{x}_{y2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ y^{(m)} & x_{m1}, & x_{m2}, & \dots, & x_{ml} & \bar{x}_{ym} \end{array} \qquad (Y)$$

Ultima coloană a celor două tabele reprezintă valoarea medie a mărimilor din linia respectivă. De exemplu

$$\bar{y}_{x1} = \frac{y_{11} + y_{12} + \dots + y_{1m}}{n_m}$$

unde n_m reprezintă numărul total de valori disponibile pe linia respectivă.

Tabelele (X) și (Y) se utilizează numai în cazul cînd se dispune de un număr mare de determinări experimentale. Dar dacă numărul datelor este destul de mic, vom utiliza tabela (A).

În cazul cînd numărul datelor experimentale este foarte mare, tabela (A) se scrie sub forma indicată mai jos în tabela (B). Aceasta se construiește în modul următor: împărțim intervalul de variație al fiecărei caracteristici X și Y în intervale egale, iar în pătrățelele astfel obținute scriem *numărul* corespunzător de valori x_k și y_k.

În prima linie a tabelei (B) sînt date valorile mărimii X care corespund mijloacelor intervalelor respective. În prima coloană sînt date valorile lui Y care de asemenea corespund mijloacelor intervalelor respective.

Scanned with OKEN Scanner

Tabela B

y \ x	x_1	x_2	x_3	...	x_m	n_m	$\bar{x}_y$
y_1	n_{11}	n_{12}	n_{13}	...	n_{1m}	n_{01}	$\bar{x}_{y1}$
y_2	n_{21}	n_{22}	n_{23}	...	n_{2m}	n_{02}	$\bar{x}_{y2}$
y_3	n_{31}	n_{32}	n_{33}	...	n_{3m}	n_{03}	$\bar{x}_{y3}$
.	.	.	.	...	.	.	.
.	.	.	.	...	.	.	.
.	.	.	.	...	.	.	.
y_s	n_{s1}	n_{s2}	n_{s3}	...	n_{sm}	n_{os}	$\bar{x}_{ys}$
n_s	n_{10}	n_{20}	n_{30}	...	n_{s0}	n \ n	
$\bar{y}_x$	$\bar{y}_{x1}$	$\bar{y}_{x2}$	$\bar{y}_{x3}$	...	$\bar{y}_{xm}$		

Penultima coloană a tabelei (B) se obţine adunînd toate elementele de pe linia corespunzătoare $(n_{10} = n_{11} + n_{12} + n_{13} \ldots + n_{1m})$, iar ultima coloană reprezintă media ponderată (vezi formula 3.3) a mărimilor x de pe linia respectivă, de exemplu,

$$\bar{x}_{y1} = \frac{1}{n_{01}} (x_1 n_{11} + x_2 n_{12} + x_3 n_{13} + \ldots + x_m n_{1m})$$

În mod analog se obţin penultima şi ultima linie.

O tabelă asemănătoare cu tabela (B) se numeşte *tabelă de corelaţie sau covarianţă.*

Deoarece numerele n_{ij} din tabela (B) se referă la mijloacele intervalelor respective, trebuie să dispunem de un număr mare de observaţii, altfel se pot introduce erori în examinarea corelaţiei dintre cele două mărimi X şi Y.

Repartiţia empirică a celor două variabile X şi Y se obţine pe cale grafică în sistemul de axe xOy, unde vom construi punctele de coordonate x_i şi y_i. Un ansamblu de puncte astfel obţinut se numeşte *cîmp de corelaţie.*

Dacă $\bar{x}$ şi $\bar{y}$ sînt mediile aritmetice respective ale tuturor valorilor obţinute pentru cele două variabile, punctul $M(\bar{x}, \bar{y})$ se va numi *centru de distribuţie.*

Tabela B este uşor de construit dacă dispunem de reprezentarea grafică a celor două variabile X şi Y deoarece avem deja cîmpul de repartiţie împărţit în pătrăţele definite de reţeaua corespunzătoare a hîrtiei milimetrice. În acest mod putem număra punctele care există într-un astfel de pătrăţel de dimensiuni stabilite de noi.

Observaţii a) Din tabela (X) se vede că luînd prima şi ultima coloană, obţinem o tabelă analogă cu tabela (A). În acest caz $\bar{y}_k$ ţine

Scanned with OKEN Scanner

locul lui Y din tabela (A), iar dacă ne-am referi la tabela (Y) ar trebui să considerăm variabile mărimile $\bar{x}_y$ în locul variabilelor X. Prin urmare, din (3.22) și (3.25) vom putea scrie

$$\left.\begin{aligned} \bar{y}_x - \bar{y} &= \frac{\mu_{11}}{\sigma_x^2}(X - \bar{x}) \\ \bar{x}_y - \bar{x} &= \frac{\mu_{11}}{\sigma_y^2}(Y - \bar{y}) \end{aligned}\right\} \tag{3.28}$$

În partea dreaptă a relațiilor (3.28), X reprezintă toate valorile $x^{(1)}, x^{(2)}, \ldots, x^{(l)}$ și, în mod similar, Y ține locul fiecăreia din valorile $y^{(1)}, y^{(2)}, \ldots, y^{(m)}$. Prin urmare, relațiile (3.28) reprezintă sisteme de ecuații, fiecare ecuație corespunzînd unei mărimi observate $x^{(i)}$ sau $y^{(j)}$.

Ecuațiile (3.23), (3.25) și (3.28) definesc *ecuațiile de regresie* iar coeficienții

$$\rho_{y/x} = \frac{\mu_{11}}{\sigma_x^2}, \qquad \rho_{x/y} = \frac{\mu_{11}}{\sigma_y^2} \tag{3.29}$$

se numesc *coeficienți de regresie.*

b) Termenul de regresie se datorește lui Galton care a aplicat pentru prima dată teoria corelației la date biologice. Examinînd dinamica populațiilor el a constatat că în astfel de cazuri s-ar păstra un echilibru dinamic dacă noile generații ar moșteni caracteristicile părinților. Cu această ocazie, studiind înălțimea taților și cea a fiilor lor, F. Galton a constatat că fiii se abat de la înălțimea medie mai puțin decît se abat tații, deci că fiii regresează spre valoarea medie. De aceea, el a folosit termenul de linie de regresie pentru linia de legătură dintre înălțimile fiilor și taților lor. Prin urmare, termenul de „regresie" nu este destul de potrivit din punct de vedere etimologic, cu toate acestea el se păstrează în literatura de specialitate deoarece s-a înrădăcinat aici și s-ar ivi prea multe complicații la o încercare de a-l înlocui cu un alt termen mai adecvat. Acest cuvînt trebuie păstrat și pe mai departe ca o simplă convenție.

c) Relațiile (3.28) arată că, în general, se pot calcula două linii de regresie pentru cazul a două variabile: i) linia de regresie a lui y spre x, adică dreapta obținută prin metoda celor mai mici pătrate în ipoteza că valorile lui x sînt exacte, și ii) linia de regresie a lui x spre y, adică dreapta găsită în ipoteza că valorile lui y sînt exacte.

3.6. Determinarea coeficientului de corelație

Să considerăm că am determinat centrul de distribuție definit de coordonatele $\bar{x}$ și $\bar{y}$, și efectuăm următoarea translație.

$$u = x - \bar{x}, \qquad v = y - \bar{y} \tag{3.30}$$

ceea ce ne permite să scriem mai departe

$$\bar{u}_v = \bar{x}_y - \bar{x}, \qquad \bar{v}_u = \bar{y}_x - \bar{y}, \tag{3.31}$$

Ţinînd seamă de (3.31) şi de (3.28), putem scrie

$$\bar{v}_u = \frac{\mu_{11}}{\sigma_x^2} u, \qquad \bar{u}_v = \frac{\mu_{11}}{\sigma_y^2} v \tag{3.32}$$

Ecuaţiile (3.32) sînt raportate la un sistem de axe cu originea în centrul de distribuţie $M(\bar{x}, \bar{y})$, astfel că din (3.30) avem mai departe

$$x = u + \bar{x}, \qquad y = v + \bar{y}. \tag{3.33}$$

Dacă însumăm în (3.33) şi împărţim la n, obţinem

$$\bar{u} = \bar{v} = 0$$

$$\sigma_x^2 = \sigma_u^2, \qquad \sigma_y^2 = \sigma_v^2. \tag{3.34}$$

Dacă scriem formulele (3.18) şi (3.5) pentru noile variabile (cu $n_k = 1$,) avem

$$\mu_{11} = \frac{1}{n}\sum_{k=1}^{n} u_k v_k, \quad \sigma_u^2 = \frac{1}{n}\sum_{k=1}^{n} u_k^2, \quad \sigma_v^2 = \frac{1}{n}\sum_{k=1}^{n} v_k^2 \tag{3.35}$$

Relaţiile (3.32) reprezintă două drepte care trec prin centrul de distribuţie $(\bar{x}, \bar{y})$, iar direcţiile acestora ne permit să determinăm coeficienţii de regresie. Primul dintre aceştia este tangenta unghiului pe care-l formează prima dreaptă de regresie (I) a lui v spre u şi axa Mu; cel de al doilea este tangenta unghiului dintre dreapta de regresie (II) a lui u spre v şi axa Mv. În figura 3.2, cele două unghiuri sînt notate prin $\hat{\alpha}$ şi $\hat{\beta}$.

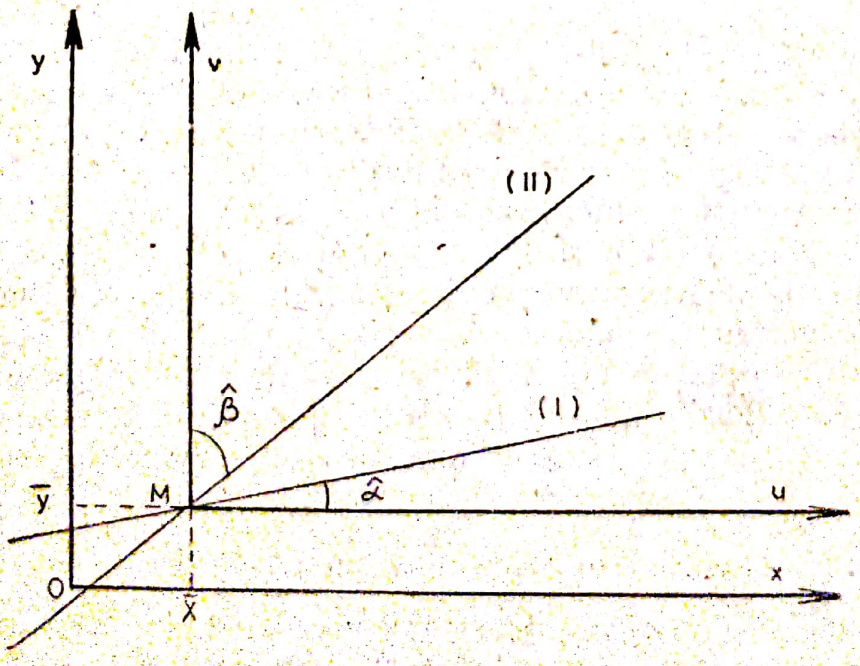


Fig. 3.2. Poziţia relativă a celor două drepte de regresie (I) şi (II).

Din relaţiile (3.29) se vede că coeficienţii de regresie trebuie să fie amîndoi pozitivi sau amîndoi negativi după cum $\mu_{11} > 0$ respectiv $\mu_{11} < 0$.

Dacă dependenţa corelaţională se transformă într-o dependenţă funcţională liniară, cele două drepte de regresie trebuie să coincidă. În acest caz trebuie să avem

$$\operatorname{tg}\hat{\alpha} \cdot \operatorname{tg}\hat{\beta} = 1 \tag{3.36}$$

deoarece

$$\hat{\alpha} + \hat{\beta} = 90^\circ$$

Dacă între u şi v *nu există legături*, atunci $\bar{v}_u$ variază puţin cînd u variază şi invers. În acest caz $\hat{\alpha}$ şi $\hat{\beta}$ sînt aproape zero, astfel că

$$\operatorname{tg}\hat{\alpha} \cdot \operatorname{tg}\hat{\beta} = 0 \tag{3.37}$$

Prin urmare, rădăcina pătrată din numărul

$$\operatorname{tg}\hat{\alpha} \cdot \operatorname{tg}\hat{\beta} = \frac{\mu_{11}^2}{\sigma_x^2 \sigma_y^2} \tag{3.38}$$

poate fi luată ca un criteriu pentru *determinarea legăturilor de corelaţie* şi „dependenţă" funcţională liniară. Astfel

$$r = \frac{\mu_{11}}{\sigma_u \cdot \sigma_v} = \frac{\frac{1}{n}\sum_{k=1}^{n} u_k v_k}{\sigma_u \cdot \sigma_v} = \frac{\sum_{k=1}^{n}(x_k - \bar{x})(y_k - \bar{y})}{n\,\sigma_x \cdot \sigma_y} \tag{3.39}$$

se va numi *coeficient de corelaţie*, în timp ce mărimea

$$\mu_{11} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})(y_k - \bar{y})$$

se numeşte covarianţă sau *corelaţie*.

Din relaţia (3.39) se vede că valoarea coeficientului de corelaţie, r, este aceeaşi indiferent dacă x sau y este luat ca variabilă independentă. De aici, şi referindu-ne la concluzia lui Galton, constatăm că dacă înălţimile fiilor regresează spre valoarea medie, atunci, dacă corelaţia dintre înălţimile taţilor şi fiilor este liniară, şi înălţimile taţilor regresează spre o aceeaşi valoare medie.

Menţionăm că acest coeficient de corelaţie este singurul dintre coeficienţii pe care-i întîlnim în teoria corelaţiei şi care se bucură de proprietatea că $r = r_{y/x} = r_{x/y}$.

Dacă ţinem seamă de expresiile dispersiilor,

$$\sigma_x^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2, \qquad \sigma_v^2 = \frac{1}{n}\sum_{k=1}^{n}(y_k - \bar{y})^2, \tag{3.40}$$

formula (3.39) se poate scrie

$$r = \frac{\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{n} (x_k - \bar{x})^2 \sum_{k=1}^{n} (y_k - \bar{y})^2}} \tag{3.41}$$

sau, dacă numărul de determinări diferă de unitate ($n_k \neq 1$), avem

$$\sigma_x^2 = \frac{1}{n}\sum_{k=1}^{n} n_k (x_k - \bar{x})^2, \quad \sigma_y^2 = \frac{1}{n}\sum_{k=1}^{n} n_k (y_k - \bar{y})^2 \tag{3.42}$$

iar pentru coeficientul de corelaţie rezultă

$$r = \frac{\sum_{k=1}^{n} n_k (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{n} n_k (x_k - \bar{x})^2 \sum_{k=1}^{n} n_k (y_k - \bar{y})^2}} \tag{3.43}$$

Să ne întoarcem acum la ecuaţiile de regresie (3.28) unde facem uz de relaţia

$$r = \frac{\mu_{11}}{\sigma_x \cdot \sigma_y}, \tag{3.44}$$

Apoi, făcînd înlocuirea $\mu_{11} = r\sigma_x\sigma_y$, din (3.28) avem

$$\left.\begin{aligned} \bar{y}_x - \bar{y} &= r\frac{\sigma_y}{\sigma_x}(X - \bar{x}) \\ \bar{x}_y - \bar{x} &= r\frac{\sigma_x}{\sigma_y}(Y - \bar{y}) \end{aligned}\right\} \tag{3.46}$$

Dacă în formula (3.41) efectuăm înmulţirile şi apoi însumăm, avem

$$\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y}) = \sum_{k=1}^{n} x_k y_k - n\bar{x}\bar{y}$$

$$\sum_{k=1}^{n} (x_k - \bar{x})^2 = \sum_{k=1}^{n} x_k^2 - n(\bar{x})^2$$

$$\sum_{k=1}^{n} (y_k - \bar{y})^2 = \sum_{k=1}^{n} y_k^2 - n(\bar{y})^2$$

iar expresia coeficientului de corelaţie devine

$$r = \frac{\sum_{k=1}^{n} x_k y_k - n\bar{x}\bar{y}}{\sqrt{\left[\sum_{k=1}^{n} x_k^2 - n(\bar{x})^2\right] \cdot \left[\sum_{k=1}^{n} y_k^2 - n(\bar{y})^2\right]}} \tag{3.41a}$$

Scanned with OKEN Scanner

Procedînd în mod analog, din formula (3.43) rezultă

$$r = \frac{\sum_{k=1}^{n} n_k x_k y_k - n\bar{x}\bar{y}}{\sqrt{\left[\sum_{k=1}^{n} n_k x_k^2 - n(\bar{x})^2\right] \cdot \left[\sum_{k=1}^{n} n_k y_k^2 - n(\bar{y})^2\right]}} \tag{3.43a}$$

unde avem

$$\bar{x} = \frac{\Sigma n_k x_k}{\Sigma n_k}, \qquad \bar{y} = \frac{\Sigma n_k y_k}{\Sigma n_k}, \qquad k = \overline{1, n}, \qquad \Sigma n_k = n$$

În anumite cazuri este mai convenabil să utilizăm formulele (3.41a) sau (3.43a) deoarece nu trebuie să calculăm diferențele $x_k - \bar{x}$ și respectiv, $y_k - \bar{y}$. În astfel de cazuri numărul operațiilor matematice se reduce considerabil, mai ales atunci cînd avem de-a face cu date experimentale foarte numeroase.

Dacă mărimile medii $\bar{x}$ și $\bar{y}$ ale celor două serii de date experimentale, x_k și y_k, sînt nule, formulele (3.41a) și (3.43a) devin

$$r = \frac{\Sigma x_k y_k}{\sqrt{\Sigma x_k^2 \cdot \Sigma y_k^2}} \tag{3.41b}$$

și respectiv,

$$r = \frac{\Sigma n_k x_k y_k}{\sqrt{\Sigma n_k x_k^2 \cdot \Sigma n_k y_k^2}} \tag{3.43b}$$

$k = \overline{1, n}$.

3.7. Proprietățile coeficientului de corelație

Dacă s_u^2 și s_v^2 reprezintă dispersiile în ecuațiile de regresie, ținînd seama de (3.32), (3.34) și (3.39), putem scrie

$$ns_v^2 = \sum_{k=1}^{n} (v_k - \bar{v}_u)^2 = \sum_{k=1}^{n} \left(v_k - \frac{\sigma_v}{\sigma_u} r \cdot u_k\right)^2$$

sau

$$ns_v^2 = \sum_{k=1}^{n} v_k^2 - 2r\frac{\sigma_v}{\sigma_u}\sum_{k=1}^{n} u_k v_k + r^2\frac{\sigma_v^2}{\sigma_u^2}\sum_{k=1}^{n} u_k^2 \tag{3.47}$$

Utilizînd notațiile

$$\sum_{k=1}^{n} v_k^2 = n\sigma_v^2, \qquad \sum_{k=1}^{n} u_k^2 = n\sigma_u^2$$

$$\sum_{k=1}^{n} u_k v_k = n\mu_{11} = nr\sigma_u\sigma_v \tag{3.48}$$

Scanned with OKEN Scanner

relația (3.47) se scrie mai departe

$$ns_v^2 = n(\sigma_v^2 - 2r^2\sigma_v^2 + r^2\sigma_i^2)$$

sau

$$s_v^2 = (1 - r)^2\sigma_v^2 \tag{3.49}$$

În mod analog putem scrie

$$s_u^2 = (1 - r)^2\sigma_u^2 \tag{3.50}$$

Din examinarea relațiilor (3.49) și (3.50) rezultă următoarele concluzii:

a) Dacă $r = \pm 1$, avem $s_v^2 = s_u^2 = 0$, iar ecuațiile de regresie (3.46) devin

$$\left.\begin{aligned} \bar{y}_x - \bar{y} &= \pm \frac{\sigma_y}{\sigma_x}(x - \bar{x}) \\ \bar{x}_y - \bar{x} &= \pm \frac{\sigma_x}{\sigma_y}(y - \bar{y}) \end{aligned}\right\} \tag{3.51}$$

Prin urmare, pentru $r = \pm 1$, între mărimile x și y există o dependență funcțională, în acest caz avem valoarea maximă a lui r și coincide cu situația cînd toate punctele observate sînt așezate pe dreapta care am determinat-o prin metoda celor mai mici pătrate (vezi figura 3.1a). Dacă $r \neq \pm 1$ între x și y nu mai există acea dependență funcțională liniară.

Dacă $r = +1$, o dată cu creșterea lui x are loc și o creștere a lui y și invers.

Dacă $r = -1$, o dată cu creșterea lui x are loc o descreștere a lui y și invers.

b) Dacă $r = 0$, între x și y nu poate exista vreo corelație definită printr-o dependență funcțională liniară. Dar aceasta nu înseamnă că între mărimile respective nu există nici un fel de corelație. În acest caz pot exista dependențe funcționale dar nu liniare. Din această cauză se spune că *r este o măsură a dependenței liniare* dintre mărimile x și y.

După cum rezultă din expresia (3.41) sau (3.43), coeficientul de corelație este o mărime abstractă și nu depinde de unitățile în care sînt exprimate variabilele x și y. Tot de aici se vede și faptul că semnul coeficientului de corelație depinde de semnul sumei $\Sigma(x_k - \bar{x})(y_k - \bar{y})$; în cazul cînd predomină abaterile perechi cu același semn, atunci r va avea semnul plus și invers, dacă predomină abaterile cu semne diferite, r va fi negativ.

Scanned with OKEN Scanner

Pentru evaluarea deviației standard a coeficientului de corelație r, se poate utiliza formula aproximativă.

$$\sigma_r = \frac{1 - r^2}{n - 1} \tag{3.52}$$

unde n reprezintă numărul de perechi de valori x_i, y_i (vezi Romanovschi, 1961, p. 386).

3.8. Corelația rangurilor

La determinarea coeficientului de corelație, pînă acum ne-am servit de perechile de mărimi x_i, y_i determinate cu o anumită precizie. Există însă cazuri în care ne interesează gradul de corelație dintre cele două serii de mărimi, dar cînd nu putem efectua măsurătorile respective. În anumite tipuri de probleme am putea avea de a face cu caracteristici care nu pot fi exprimate prin cifre, dar la care ne interesează totuși existența unui anumit grad de corelare. Ca un exemplu de acest fel ar putea fi dată relația dintre aptitudinile matematice și cele muzicale la elevii dintr-o anumită clasă. De la un elev la altul putem întîlni aptitudini diferite, iar aprecierea acestora prin date numerice poate fi supusă la o serie de obiecții și depinde de modul de lucru al examinatorilor precum și metodele utilizate de ei. De aceea, gradul de corelație pentru notele de la matematică și cele de la muzică nu va oglindi în mod destul de obiectiv relația dintre cele două aptitudini — la matematică și la muzică.

În astfel de cazuri, dificultățile amintite pot fi evitate întrucîtva dacă ne vom referi mai mult la ordinea aptitudinilor și nu la evaluarea numerică a acestor parametri. În modul acesta, fiecărei persoane examinate îi vom atribui un număr de ordine care va indica poziția celui examinat în seria persoanelor de care ne interesăm; astfel vom nota de exemplu pe cel mai bun cu cifra 1, apoi următorul cu 2 și așa mai departe. Procedînd astfel, cei examinați vor fi *ordonați*, aranjați într-o anumită ordine, iar numărul pe care l-am atribuit unei anumite persoane se numește *rang*.

Un alt exemplu de acest fel ar putea fi legat de influența pe care corpurile cerești ar avea-o asupra organismului uman. În prezent nu dispunem de mijloace tehnice pentru măsurarea precisă asupra gradului unei astfel de influențe, de aceea s-ar putea utiliza doar ordonarea corpurilor respective după anumiți parametri cunoscuți; masă, densitate, temperatura de la suprafață, compoziția chimică a atmosferelor respective, depărtarea de la Pămînt etc.

Alte exemple de utilizare a rangurilor pot fi amintite: ordonarea unui număr de oameni după înălțime, fără a efectua măsurători; aranjarea unui număr de nuanțe de culori; ordonarea unui număr de

țări după importanța producției anumitor bunuri sau după venitul național pe cap de locuitor etc, cazuri în care măsurarea numerică efectivă ar fi destul de dificilă sau imposibilă.

În cazul cînd rangul anumitor caracteristici este mai accesibil decît măsurările efective, pentru examinarea legăturilor dintre anumiți parametri, putem cerceta corelația rangurilor respective. În acest caz metodele parametrice amintite mai înainte se înlocuiesc cu metode neparametrice care ne vor permite să determinăm intensitatea corelației care există între anumite caracteristici cantitative sau calitative. Prezentăm mai jos cîteva procedee de evaluare a anumitor corelații în care se ține seama de ordonarea mărimilor caracteristice, adică în care se utilizează rangul sau poziția pe care le au anumite mărimi, una în raport cu alta. Vom determina, deci, coeficientul de corelație al rangurilor.

Fie n numărul de elemente ale unei mulțimi ordonate după două caracteristici calitativ diferite; pentru caracteristica A să dispunem de ordonarea: $x_1, x_2, \ldots, x_n$ iar pentru caracteristica B să avem: $y_1, y_2, \ldots, y_n$. unde x și y sînt simple permutări ale celor n numere naturale.

Să notăm aici

$$d_k = x_k - y_k$$

unde vectorul de componente d_k $(k = \overline{1,n})$ este o mărime a gradului de apropiere pe care o poate avea corespondența dintre A și B. Dacă $d = 0$, înseamnă că toți indivizii considerați au cele două caracteristici în aceeași ordine, avem o corespondență perfectă, rangul x_k al caracteristicii A este același ca și rangul y_k al caracteristicii B. În realitate s-ar putea ca $x_k = y_k$, adică, în exemplul cu clasa de elevi, un elev să fie primul atît la matematică cît și la muzică, dar dacă $x_k \neq y_k$, atunci am putea avea un elev printre primii la matematică și printre ultimii la muzică sau invers.

Pentru o cercetare a corelației rangurilor, vom considera scrise în ordinea crescătoare sau descrescătoare toate valorile lui x iar alături de acestea scriem valorile corespunzătoare ale lui y, acestea din urmă putînd fi sau nu în vreo anumită ordine. Prin urmare, în cadrul ordonării după rang, x ia valorile: 1, 2, 3, ..., n în anumită ordine, pe cînd y ia aceleași numere dar care pot să nu respecte nici o ordine. Din aceste considerații se vede imediat că valorile medii $\bar{x}$ și $\bar{y}$ se pot înlocui prin media rangurilor care atît pentru x cît și pentru y va fi dată de relația

$$\bar{x} = \bar{y} = \frac{n+1}{2}, \qquad (3.53)$$

adică media celor două caracteristici este înlocuită prin media rangurilor corespunzătoare.

Scanned with OKEN Scanner

Pentru definirea coeficientului corelației rangurilor, vom scrie

$$\sum_{k=1}^{n}(x_k-\bar{x})^2=\sum_{k=1}^{n}(y_k-\bar{y})^2=\sum_{k=1}^{n}(y_k^2-2\bar{y}y_k+\bar{y}^2)$$

sau, ținînd seama de (3.53), avem

$$\sum_{k=1}^{n}(x_k-\bar{x})^2=\sum_{k=1}^{n}(y_k-\bar{y})^2=\frac{n(2n+1)(n+1)}{6}-(n+1)n\frac{n+1}{2}+n\left(\frac{n+1}{2}\right)^2$$

adică

$$\sum_{k=1}^{n}(x_k-\bar{x})^2=\sum_{k=1}^{n}(y_k-\bar{y})^2=\frac{n(n^2-1)}{12} \tag{3.54}$$

În continuare ne vom folosi de faptul că putem scrie următoarea identitate

$$\sum_{k=1}^{n}(x_k-\bar{x})(y_k-\bar{y})=\frac{1}{2}\left[\sum_{k=1}^{n}(x_k-\bar{x})^2+\sum_{k=1}^{n}(y_k-\bar{y})^2\right]-$$

$$-\frac{1}{2}\sum_{k=1}^{n}[(x_k-\bar{x})-(y_k-\bar{y})]^2 \tag{3.55}$$

Valabilitatea relației (3.55) se verifică cu ușurință efectuînd operațiile indicate în termenii din partea dreaptă și păstrînd partea stîngă nemodificată.

În relația (3.55), primul termen din partea dreaptă este calculat cu formula (3.54), iar cel de-al doilea termen va fi notat prin

$$-\frac{1}{2}\sum_{k=1}^{n}[(x_k-\bar{x})-(y_k-\bar{y})]^2=-\frac{1}{2}\sum_{k=1}^{n}d_k^2$$

unde se ține seama de (3.53), adică $\bar{x}=\bar{y}$. Prin urmare,

$$d_k=(x_k-\bar{x})-(y_k-\bar{y})=x_k-y_k \tag{3.56}$$

În modul acesta, relația (3.55) se poate scrie

$$\sum_{k=1}^{n}(x_k-\bar{x})(y_k-\bar{y})=\frac{n(n^2-1)}{12}=\frac{1}{2}\sum_{k=1}^{n}d_k^2 \tag{3.57}$$

unde, așa cum am menționat mai sus, d_k reprezintă diferența de ranguri x_k-y_k pentru variabila x ordonată într-un anumit sens — *crescător sau descrescător,* iar rangurile lui y sînt acele care corespund acestei mărimi pentru care caracteristica x este x_k.

Dacă θ va reprezenta coeficientul de corelație al rangurilor, atunci înlocuind pe (3.57) și (3.54) în formula (3.41) corespun-

Scanned with OKEN Scanner

zătoare pentru coeficientul de corelație al mărimilor caracteristice, avem

$$\theta = \frac{n\frac{n^2-1}{12} - \frac{1}{2}\sum_{k=1}^{n} d_k^2}{\sqrt{n^2 \frac{n^2-1}{12} \cdot \frac{n^2-1}{12} n}} = \frac{n(n^2-1) - 6\sum d_k^2}{n(n^2-1)}$$

sau

$$\theta = 1 - \frac{6\sum_{k=1}^{n} d_k^2}{n(n^2-1)} \tag{3.58}$$

care se cunoaște sub denumirea de *coeficient al lui Ch.E.Spearman*. *Observații*: a) Pentru două serii de ranguri x și y identice, adică atunci cînd pentru același cuplu de valori x și y avem aceleași rang, $d_k = y_k - y_k = 0$, iar din (3.58) rezultă $\theta = 1$.

b) Dacă cele două serii de valori x și y au rangurile exact invers, de exemplu pentru x avem rangurile: 1, 2, 3, ..., n, iar pentru mărimile y corespunzătoare avem rangurile: n, $n-1$, $n-2$, ..., 1, atunci, pentru valorile lui d, putem scrie

$$d = 1-n,\ 2-(n-1),\ \ldots,\ \left\{\begin{matrix} -1, & +1, \\ -2, & 0, \quad +2 \end{matrix}\right\},\ \ldots,\ (n-2),\ n-1$$

Valorile din prima linie din acoladă corespund cazului cînd n este par ($n = 2p$) iar valorile din linia a doua corespund cazului cînd n este impar ($n = 2p+1$), unde n reprezintă numărul de termeni, adică numărul perechilor de valori x_k, y_k.

În cazul cînd n este par ($n = 2p$), pentru $\sum_{k=1}^{n} d_k^2$ putem scrie

$$\sum_{k=1}^{n} d^2 = [1-2p]^2 + [2-(2p-1)]^2 + [3-(2p-2)]^2 + \ldots + [(2p-1)-2]^2 + [2p-1]^2$$

Șirul acesta de termeni care se adună în expresia lui $\sum_{k=1}^{n} d_k^2$ este simetric față de intervalul dintre termenii de la mijloc. Prin urmare putem considera că $\sum_{k=1}^{n} d_k^2$ va fi egală cu dublul sumei primilor $n/2 = p$ termeni, unde primul termen este $(1-2p)^2$ sau $(2p-1)^2$, ultimul termen al acestui șir este $[p-(2p-(p-1))]^2 = 1^2$ iar penultimul termen este $[p-1-(2p-(p-2))]^2 = 3^2$ etc.

Pe baza acestor considerații, pentru $n = 2p$, avem

$$\sum_{k=1}^{n} d_k^2 = 2[1^2 + 3^2 + \ldots + (2p-1)^2] = \frac{2}{3} p(2p+1)(2p-1) \qquad (3.59)$$

unde ne-am folosit de formula care dă suma pătratelor primelor n numere naturale impare

$$S_1 = 1^2 + 3^2 + \ldots + (2p-1) = \frac{1}{3} \cdot p(2p+1)(2p-1).$$

În cazul cînd n este număr impar ($n = 2p + 1$), avem

$$\sum_{k=1}^{n} d_k^2 = [1 - (2p+1)]^2 + [2 - 2p]^2 + [3 - (2p-1)]^2 + \ldots +$$

$$+ [(2p-1) - 3]^2 + [2p-2]^2 + [(2p+1) - 1]^2.$$

Termenii acestei sume sînt simetrici față de termenul de la mijloc. Primul termen va fi $(2p)^2$, al doilea este $(2 - 2p)^2$ etc. Prin urmare, suma $\sum_{k=1}^{n} d_k^2$ va fi egală cu dublul produsului sumei pătratelor numerelor pare pe care le conține, adică

$$\sum_{k=1}^{n} d_k^2 = 2[2^2 + 4^2 + \ldots + (2p)^2] = \frac{4}{3} p(p+1)(2p+1) \qquad (3.60)$$

unde ne-am folosit de formula care dă suma pătratelor primelor n numere naturale pare

$$S_2 = 2^2 + 4^2 + \ldots + (2p)^2 = \frac{2}{3} p(p+1)(2p+1)$$

Acum este ușor de văzut că dacă înlocuim pe (3.59) în (3.58) obținem $\theta = -1$. Același lucru se întîmplă și atunci cînd înlocuim pe (3.60) în (3.58). Prin urmare

$$-1 \leqslant \theta \leqslant +1$$

Reamintim că pentru $\theta = +1$, caracteristicile x_k au același rang ca și caracteristicile y_k, în timp ce pentru $\theta = -1$, ordonarea lui y este exact inversă celei a lui x.

Un alt coeficient utilizat pentru corelația rangurilor este așa-numitul coeficient a lui M.Kendall. Pentru definirea acestuia se întocmește o tabelă cu cele două serii de ranguri R_x și R_y (vezi coloanele 4 și 5 din tabela 3.1) unde elementele respective sînt ordonate, de exemplu, după valorile crescătoare ale lui R_x, în coloana alăturată acesteia fiind scrise mărimile R_y corespunzătoare.

Scanned with OKEN Scanner

Între mărimile R_x și R_y se pot stabili perechi de valori, numărul maxim al unor astfel de combinații posibile fiind definit prin relația

$$C_n^2 = \frac{n(n-1)}{2}$$

(C_n^2 = combinări de n elemente luate cîte două)

Dar dacă fixăm ordonarea după valorile crescătoare ale rangurilor lui x, rangurile lui y fiind oarecare, înseamnă că putem combina un rang al lui y numai cu rangurile care, în secvența respectivă, urmează după el în sensul crescător al lui R_x. Prin urmare, vom putea face atîtea combinații ale rangurilor R_x și R_y cît este suma algebrică a rangurilor superioare luate cu semnul pozitiv, Q, plus rangurile inferioare luate cu semn negativ, P, și care urmează după un anumit rang R_y considerat. În tabela 3.1, pentru primul rang al lui R_y avem scris numărul 12 ceea ce înseamnă că după acesta, în sensul crescător al lui R_x, avem rangurile mai mari $R_y = 13, 14, 16, 17, 15$ (vezi coloana 5 din tabela 3.1) adică $Q = 5$, în timp ce rangurile mai mici sînt $R_y = 10, 11, 8, 9, 6, 2, 4, 3, 7, 5, 1$ al căror număr luat cu semnul minus ne conduce la $P = -11$.

În modul acesta, alături de coloanele rangurilor se mai scriu două coloane: una cu rangurile negative și mai mici, P, iar alta cu rangurile mai mari și pozitive Q. Apoi, suma algebrică $S_i = P_i + Q_i$ se află în coloana următoare, iar suma $S = \Sigma S_i$ ne dă numărul total de combinații care se pot alcătui în ordonarea considerată. Prin urmare, raportul

$$\tau = \frac{2S}{n(n-1)} \tag{3.61}$$

este un coeficient de corelație a rangurilor. Acesta a fost introdus de către Kendall și are sensul unei probabilități empirice

$$\tau = \frac{\text{Număr de puncte din ordonările considerate}}{\text{Număr total de puncte posibile în ordonări}}$$

Coeficientul lui Kendall, τ, poate avea valoare $+1$ cînd două serii de valori au ranguri indentice, deoarece în acest caz vom avea numai ranguri superioare, iar suma acestora va fi

$$(n-1) + (n-2) + \ldots + 2 + 1 = \frac{n(n-1)}{2}$$

În cazul cînd $\tau = -1$, avem ordonarea lui y exact invers ordonării lui x; în acest caz pentru fiecare rang al lui y, în ordinea considerată, avem numai ranguri inferioare ale căror sumă generală este $-n(n-1)/2$.

Scanned with OKEN Scanner

Tabela 3.1

Țara	Venit național pe un locuitor în 1965 (dolari) x	Născuți la 1000 locuitori în 1965 y	Rangul după Venit național R_x	Născuți vii la 1000 locuitori R_y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$d = R_x - R_y$	d^2	P	Q	$S = P+Q$
S.U.A.	2893	18,5	1	12	2045,6	−11,2	− 22910,7	4184479,36	125,44	−11	121	−11	5	− 6
Canada	1825	19,5	2	10	977,6	−10,2	− 9971,5	955701,76	104,04	− 8	64	− 9	6	− 3
Anglia	1451	17,9	3	13	603,6	−11,8	− 7122,5	364332,96	139,24	−10	100	−10	4	− 6
R.F.G.	1447	17,8	4	14	599,6	−11,9	− 7135,2	359520,16	141,61	−10	100	−10	3	− 7
Franța	1436	17,5	5	16	588,6	−12,2	− 7180,9	346449,96	148,84	−11	121	−11	1	−10
Belgia	1406	15,8	6	17	558,6	−13,9	− 7764,5	312033,96	193,21	−11	121	−11	0	−11
Austria	970	17,6	7	15	122,6	−12,1	− 1483,5	15030,76	146,41	− 8	64	−10	0	−10
Italia	883	18,9	8	11	35,6	−10,8	− 384,5	1267,36	116,64	− 3	9	− 9	0	− 9
Argentina	740	23,0	9	8	−107,4	− 6,7	+ 719,6	11534,76	44,89	1	1	− 7	1	− 6
Portugalia	351	22,3	10	9	−496,4	− 7,4	+ 3673,4	246412,96	54,76	1	1	− 7	0	− 7
Brazilia	217	43,0	11	6	−630,4	13,3	− 8384,3	397404,16	176,89	5	25	− 5	1	− 4
Iran	211	48,0	12	2	−636,4	18,3	− 11646,1	405004,96	334,89	10	100	− 1	4	+ 3
Tunisia	179	45,1	13	4	−668,4	15,4	− 10293,4	446758,56	237,16	9	81	− 2	2	0
Maroc	174	46,1	14	3	−673,4	16,4	− 11043,8	453467,56	268,96	11	121	− 1	2	+ 1
India	86	38,4	15	7	−761,4	8,7	− 6624,2	579729,96	75,69	8	64	− 2	0	− 2
Haiti	80	45,0	16	5	−767,4	15,3	− 11741,2	588902,76	234,09	11	121	− 1	0	− 1
Birmania	56	50,0	17	1	−791,4	20,3	− 16865,4	626313,96	412,09	16	256	0	0	0
	14405	504,4					−135358,7	10294345,92	2954,85		1470			−78

r_{yx}

$n = 17$
$\bar{x} = 847,4$
$\bar{y} = 29,7$

$$r_{yx} = \frac{-135358,7}{\sqrt{10294346 \times 2955}} = -0,77608$$

$$Q = 1 - \frac{6 \times 1470}{17(17^2 - 1)} = -0,802$$

P = ranguri mai mici −
Q = ranguri mai mari +
} din coloana R_y.

$$\tau = \frac{2(-78)}{17 \times 16} = -0,574$$

Scanned with OKEN Scanner

Aplicație numerică

Pentru ilustrarea aplicației practice a considerațiilor expuse mai sus, am utilizat tabela nr. 14 din lucrarea lui Sileţchi și Curta (1972) unde se analizează relația dintre venitul național exprimat în dolari pe cap de locuitor în anul 1965 și născuții vii la 1000 locuitori, în cîteva țări.

Datele numerice corespunzătoare sînt cuprinse în primele trei coloane ale tabelei 3.1 iar semnificația celorlalte coloane ajutătoare este ușor de înțeles din notația folosită la fiecare cap de coloană. După efectuarea tuturor calculelor se obține (vezi tabela 3.1):

— Coeficientul de corelație $r = -0{,}77$,
— Coeficientul lui Spearman $0 = -0{,}80$,
— Coeficientul lui Kendall $\tau = -0{,}57$.

De aici rezultă că toți cei trei coeficienți de corelație sînt negativi, fapt care pune în evidență existența unei relații statistice inverse între cele două mărimi analizate. Dintre aceștia, coeficientul lui Kendall indică o corelație mai slabă.

Observație: Coeficienții de corelație ai rangurilor pot fi utilizați și atunci cînd nu avem date care să se supună unei repartiții normale.

3.9. Corelația seriilor dinamice

În paragrafele precedente am considerat că pentru un element dat al unei mulțimi am determinat două caracteristici și am examinat corelația dintre ele, în acest scop referindu-ne la mai multe elemente ale mulțimii, fără a specifica momentul determinărilor respective. Cu alte cuvinte, ne-am ocupat de caracteristicile pe care le-am considerat „constante" în timp, adică am considerat că avem de-a face cu semnale staționare. Dar, în anumite domenii de activitate, cele două caracteristici trebuie considerate variabile în timp, astfel de serii de date experimentale se vor numi *serii dinamice*. Prin urmare, o serie dinamică se obține atunci cînd caracteristicile respective sînt variabile în timp, iar noi efectuăm determinarea lor la momente diferite. Ca exemplu de astfel de caracteristici amintim valorile temperaturii măsurate într-un loc oarecare și la momente diferite, populația unei țări de-a lungul unui număr mare de ani, greutatea unui animal la diferite etape din viața lui etc.

Seriile de date statistice care se referă la desfășurarea în timp a fenomenelor economice și sociale, sau chiar a unor fenomene naturale, ne interesează nu numai din punctul de vedere al valorii lor ci și cît privește *ordinea* de apariție. Aceste date în general pot fi supuse unei anumite tendințe (trend), fapt care poate complica foarte mult cercetarea corelației dintre diferite caracteristici ale mai multor elemente ale unei colectivități.

Scanned with OKEN Scanner

Pentru ușurarea expunerii considerațiilor privind corelația seriilor dinamice, vom presupune că datele experimentale — observaționale sau de măsurare — sînt echidistante în timp. În modul acesta, egalînd intervalul de timp cu unitatea ($\delta t = 1$) putem considera seria de caracteristici ca fiind obținută la momentele: $t = 1, 2, 3, \ldots$, iar una dintre caracteristici va avea valorile: $x_1, x_2, x_3, \ldots$.

Caracteristica x poate fi discretă (număr de oameni) sau continuă (ploaia, temperatura etc), funcție de o singură variabilă sau mai multe variabile. Aici am putea avea *serii bidimensionale* cu două serii de caracteristici și să exprimăm corelarea dintre ele: producția de grîu și cea de porumb pe care o țară a avut-o în mai mulți ani consecutivi sau am putea considera o singură caracteristică drept o serie de valori în timp și să examinăm corelația dintre valorile acesteia și momentele corespunzătoare — *serie dinamică unidimensională.*

În figura 3.3 este dat un exemplu de serie unidimensională (vezi tabelul 3.8 a) care reprezintă dinamica producției de grîu și secară a României între anii 1951 și 1975 (vezi Mihoc, Bergthaller și Urseanu, 1978, pg.419). După cum se observă, variațiile sînt considerabile de la an la an. Totuși se poate vedea că această serie dinamică se compune din trei elemente: o variație de lungă durată numită tendință (trend) care constă în aceea că dinamica producției de grîu și secară este în creștere continuă; apoi există variații ritmice care par să se repete după un anumit interval de timp numit perioadă și, în fine, există variații neregulate care se produc în mod accidental.

Faptul că într-o anumită serie dinamică este prezentă tendința -trendul, va complica foarte mult problema privind examinarea core-

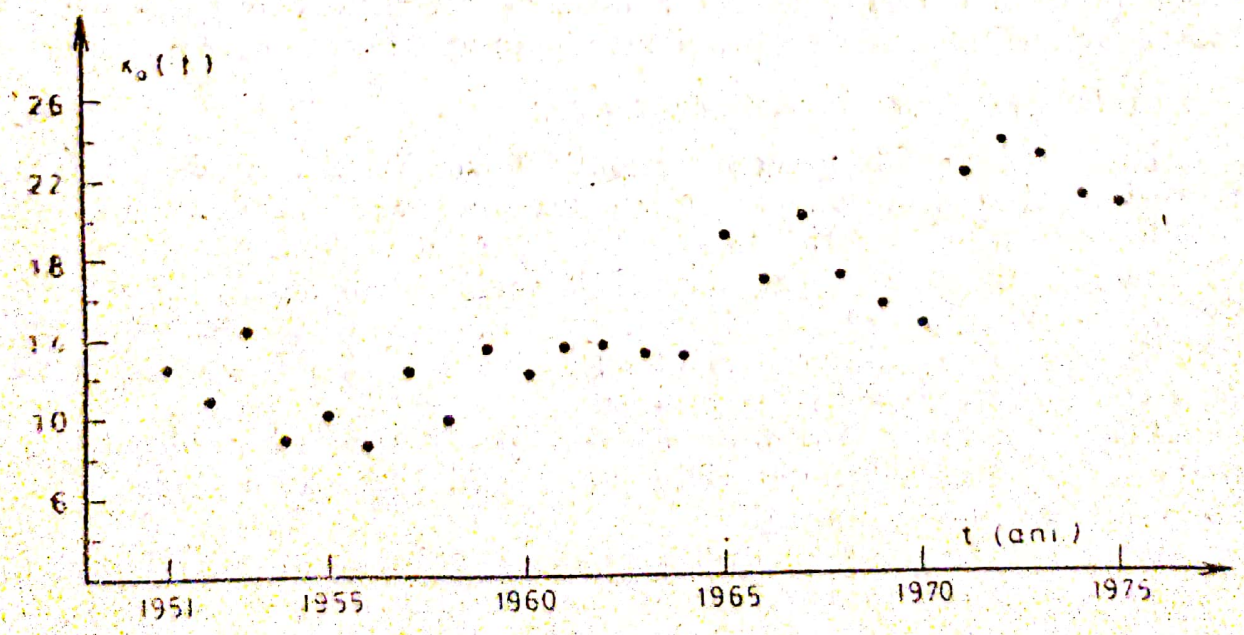


Fig. 3.3. Dinamica producției medii de grîu și secară la hectar în România între anii 1951 și 1975.

lației, de aceea va trebui mai întîi să eliminăm tendința și eventual apoi să se aplice procedeele cunoscute de calcul. În acest scop s-au făcut multe propuneri printre care amintim: studierea corelației nu între mărimile direct observate ale parametrilor respectivi ci corelația diferențelor dintre valorile consecutive ale unei serii dinamice, sau corelarea componentelor reziduale, adică diferențele față de valoarea medie corespunzătoare. În astfel de cazuri putem să reluăm problema amintită anterior — determinarea coeficientului de corelație pentru serii dinamice unidimensionale. Probleme deosebite se ridică atunci cînd avem de-a face cu serii bidimensionale unde va trebui să distingem *corelația sincronă* și cea *asincronă*.

Dacă dispunem de o serie dinamică bidimensională în care x, y sînt determinate pentru momentul t_k, atunci pe același grafic putem reprezenta cele două curbe $x = f(t)$ și $y = g(t)$. Dacă curbele astfel obținute au extremele situate aproximativ la aceeași abscisă, t_{max} sau t_{min}, atunci spunem că cele două serii dinamice sînt *sincrone*. În caz contrar trebuie să determinăm *faza de corelație*, adică faza de abatere de la sincronism — în astfel de situații seriile dinamice respective sînt considerate *asincrone*.

În anumite lucrări de statistică, *corelația sincronă* este definită ca o corelație între valorile luate de două variabile statistice distincte (x, y) la un același moment, pe cînd *corelația asincronă* sau *corelația cu decalaj de timp* este considerată corelația dintre valoarea luată de o variabilă statistică (x) la un anumit moment t și valoarea unei alte variabile (y) la momentul $t + \Delta t$ (Δt = constant).

Practic, cele două moduri de definire a corelațiilor sincrone și asincrone sînt echivalente, dar felul în care am definit corelația asincronă pune în evidență prezența fazei de corelație care poate deveni foarte importantă pentru probleme de prognoză.

a) *Determinarea fazei de corelație*

Din considerațiile prezentate mai sus vedem că faza de corelație reprezintă timpul cu care ar trebui micșorate sau mărite momentele — abscisele dintr-o serie dinamică — pentru ca extremele acesteia să coincidă cît mai bine posibil cu extremele celeilalte serii dinamice. Cu alte cuvinte, faza de corelație definește corecția de timp care trebuie aplicată la abscisele unei serii dinamice pentru ca cele două serii considerate să fie în sincronism.

Pentru determinarea fazei de corelație vom distinge două procedee, după cum mărimile celor două serii sînt sau nu exprimate în același sistem de unități.

i) În cazul cînd s-au utilizat sisteme diferite de unități putem calcula coeficientul de corelație al celor două serii $x_k = f(t_k)$ și $y_k = g(t_k)$. În acest scop, în funcția $f(t_k)$, de exemplu, vom menține fixe valorile lui t_k, iar în $g(t_k)$ lui t_k să-i aplicăm diferite corecții Δt_i.

Apoi, determinînd diferite valori pentru coeficientul de corelație, vom căuta valoarea extremă r_{max}. a acestuia. Pentru $r_i = r_{max}$ va rezulta $(\Delta t_i)_{max} = \varphi =$ faza de corelație.

Procedeul indicat mai sus pentru determinarea fazei de corelație necesită un număr mare de calcule pentru coeficientul de corelație, astfel încît curba $r_i = F(\Delta t_i)$ să fie bine definită, iar mărimea r_{max} să se poată determina cu precizie suficientă. De aceea, vom prezenta aici un sistem de aproximații succesive prin care să reducem volumul de calcul.

Considerînd că mărimile x_k, y_k sînt obținute la intervale egale de timp, sau dacă printr-un procedeu de interpolare am reușit să îndeplinim această condiție, atunci alegem în mod arbitrar $\Delta t_0 = N \cdot \delta t$, unde δt este pasul absciselor t_k iar N este un număr întreg. Constanta Δt_0 trebuie aleasă astfel încît aplicînd-o la abscisele unei serii, extremele celor două serii x_k și y_k să coincidă cît mai bine posibil.

Fie: r_0 valoarea coeficientului de corelație cînd se aplică corecția $\Delta t = N \cdot \delta t = \Delta t_0$.

r_+ valoarea coeficientului de corelație cînd se aplică corecția $\Delta t = (N + 1)\ \delta t = \Delta t_0 + \delta t$,

r_- valoarea coeficientului de corelație cînd se aplică corecția $\Delta t = (N - 1)\ \delta t = \Delta t_0 - \delta t$.

Considerînd că, pe un interval destul de mic, curba empirică se poate aproxima printr-un polinom de gradul doi, putem scrie

$$r = a(\Delta t)^2 + b \cdot \Delta t + c \tag{3.62}$$

astfel că,

$$\begin{aligned} r_- &= a[(N-1)\delta t]^2 + b[(N-1)\delta t] + c \\ r_0 &= a[N \cdot \delta t]^2 + b[N \cdot \delta t] + c \\ r_+ &= a[(N+1)\delta t]^2 + b[(N+1)\delta t] + c \end{aligned} \tag{3.63}$$

care este un sistem de trei ecuații cu cele trei necunoscute a, b și c pentru care obținem

$$a = \frac{r_+ + r_- - 2r_0}{2(\delta t)^2}, \qquad b = \frac{r_+ - r_- - 2(r_+ + r_- - 2r_0)N}{2 \cdot \delta t} \tag{3.64}$$

$$c = r_0 - a[N \cdot \delta t]^2 + b \cdot N \cdot \delta t$$

Pentru $N = 0$ avem mai departe

$$a = \frac{r_+ + r_- - 2r_0}{2(\delta t)^2}, \qquad b = \frac{r_+ - r_-}{2 \cdot \delta t}, \qquad c = r_0 \tag{3.65}$$

Scanned with OKEN Scanner

Cînd r atinge valoarea extremă ($r = r_{max}$), se anulează derivata de ordinul întîi a funcției (3.62), astfel că putem scrie

$$\varphi = (\Delta t)_{max} = -\frac{b}{2a}$$

sau, dacă ținem seama de relațiile (3.64),

$$\varphi = -\frac{2(\delta t)^2}{2(r_+ + r_- - 2r_0)} \cdot \frac{r_+ - r_- - 2(r_+ + r_- - 2r_0)N}{2(\delta t)}$$

unde folosim notația $\Delta t_0 = N \cdot \delta t$, astfel că după ce simplificăm, avem

$$\varphi = \Delta t_0 - \frac{\delta t}{2} \cdot \frac{r_+ - r_-}{r_+ + r_- - 2r_0} \qquad (3.66)$$

Aplicație practică

Pentru a ilustra aplicarea practică a considerațiilor de mai sus, am utilizat datele numerice care au rezultat din prelucrarea materialului observațional obținut pentru steaua variabilă VZ Cancri. Fără să mai intrăm în amănunte de specialitate, precizăm că pentru această stea am dispus de datele numerice care sînt reproduse în primele trei coloane ale tabelei 3.2, unde t_k sînt momentele observațiilor respective și exprimă fazele în ciclul perioadei secundare. Această noțiune va fi analizată în capitolul VI unde ne vom referi și la fenomene naturale multiperiodice.

Mărimile $x = m_{max}$ reprezintă magnitudinea stelară pe care o are steaua *VZ Cancri* în maximul de strălucire, iar cantitățile $y = O - C$ sînt diferențele dintre momentele observate și cele calculate ale maximelor respective. Și aspura acestei noțiuni vom reveni.

Prin urmare, pornim de la cele trei coloane ale tabelei 3.2 și considerăm că dispunem de cele două serii dinamice $x_k = (m_{max})_k$ și $y_k = (O - C)_k$, pentru fiecare t_k.

Graficele celor două funcții $x_k = f(t_k)$ și $y_k = g(t_k)$ sînt reproduse pe *figura 3.4* de unde se poate evalua constanta $\Delta t_0 = -0^P,250$, iar P este perioada ciclului secundar. În modul acesta putem menține fixe abscisele seriei $y_k = g(t_k)$ și să le modificăm pe acele care definesc funcția $x_k = f(t_k) = m_{max}$.

Deoarece, așa cum rezultă din tabela 3.2, avem $\delta t = 0^P,025$, din relația $\Delta t_0 = N \cdot \delta t$, rezultă $N = 10$, $N - 1 = 9$ și $N + 1 = 11$.

În coloana a 4-a a tabelei 3.2 se găsesc magnitudinile $x = m_{max}$ ale căror abscise t_k au fost corectate cu cantitatea $\Delta t_0 = N \cdot \delta t = -0^P,250$, adică magnitudinile care au abscisele $t_k - \Delta t_0$.

Astfel, de exemplu, pentru $t_k = 0^P,000 = 1^P000$ vom avea $t_k - \Delta t_0 = 1^P,000 - 0^P,250 = 0^P,750$, adică la $t_k = 0^P000$ aducem magnitudinea $x = m_{max} = 7,30$ care se află, în coloana a 2-a, la $t_k = 0^P,750$. La

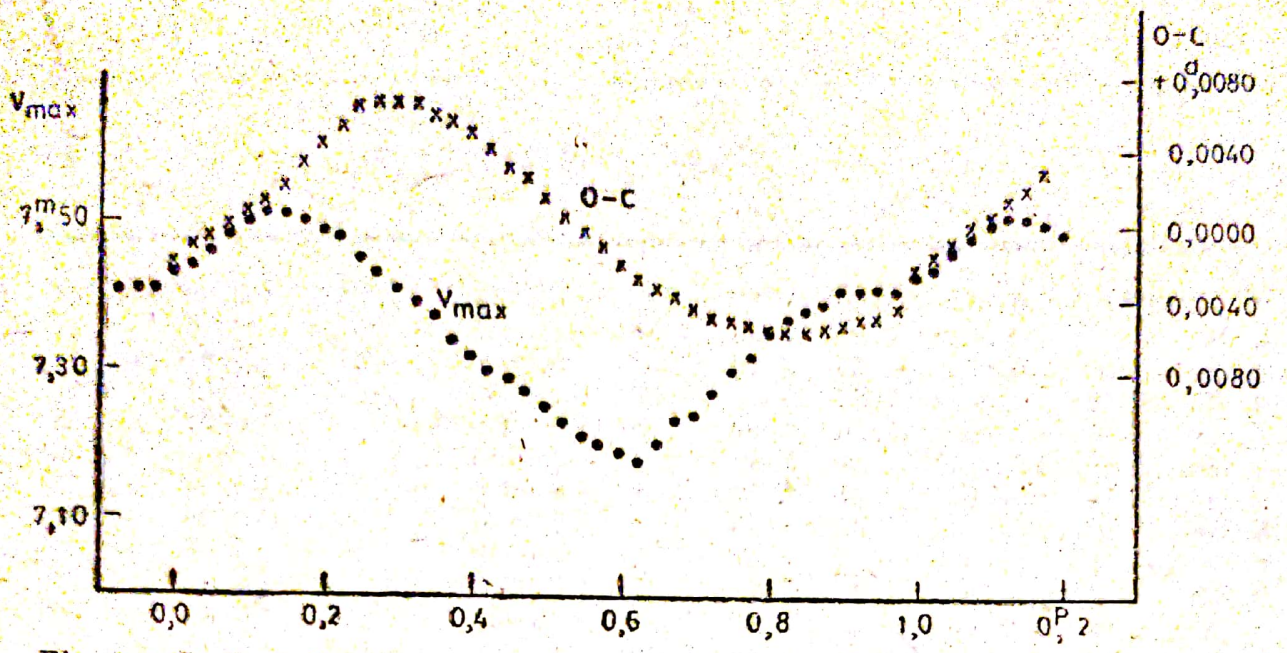


Fig. 3.4. Graficul datelor numerice pentru faza de corelație la steaua VZ Cancri.

$t_k = 0^P,250$ aducem magnitudinea $x = m_{max} = 7,43$ care, în coloana a 2-a, se găsea la $t_k = 0^P,000$ etc.

Coloana a 5-a se obține în modul indicat mai sus, aplicînd corecția $\Delta t_- = \Delta t_0 - \delta t = -0^P,250 - 0^P,025 = -0^P,275$. Astfel, de exemplu, magnitudinea $x = m_{max} = 7^P,27$ care se găsea la $t_k = 0^P,725$ va fi deplasată în dreptul lui $t_k = 0^P,000$, iar $x = m_{max} = 7,30$ va fi deplasată de la $t_k = 0^P,750$ la $t_k = 0^P,025$ etc.

Coloana a 6-a se obține aplicînd corecția $\Delta t_+ = -0^P,250 + 0^P,025 = -0^P,225$. Astfel $x = m_{max} = 7,32$ de la $t_k = 0^P,775$ va fi trecută la $t_k = 0^P,000$ iar $x = m_{max} = 7,30$ de la $t_k = 0^P,750$ va fi trecută la $t_k = 0^P,975$.

În continuare, la capul fiecărei coloane am specificat conținutul corespunzător.

Avînd în vedere permutările care se fac în alcătuirea diferitor coloane, putem scrie

$$\Sigma(x - \bar{x})_0^2 = \Sigma(x - \bar{x})_-^2 = \Sigma(x - \bar{x})_+^2$$

ceea ce înseamnă că atunci cînd aplicăm formula (3.41) pentru calculul coeficienților de corelație, r_+, r_0, și r_-, valoarea numitorului rămîne nemodificată.

Rezumînd datele din tabela 3.2, putem scrie:

$$\Delta t_0 = -0^P,250, \quad \delta t = 0^P,025, \quad \Sigma(y_k - \bar{y})^2 = 0,00075255,$$
$$\Sigma(x_k - \bar{x})_0^2 = 0,4064, \quad \Sigma(y_k - \bar{y})(x_k - \bar{x})_0 = 0,016008,$$
$$\Sigma(y_k - \bar{y})(x_k - \bar{x})_- = 0,015150, \quad \Sigma(y_k - \bar{y})(x_k - \bar{x})_+ = 0^P,016567$$

Tabela 3.2

t	x	y	x			$y-\bar{y}$	$(x-\bar{x})_0$	$(x-\bar{x})_-$	$(x-\bar{x})_+$
			Δt_0	Δt_-	Δt_+				
$0^p,000$	$7^m,43$	$-0^d,0024$	$7^m,30$	$7^m,27$	$7^m,32$	$-0^d,0022$	$-0^m,06$	$-0^m,09$	$-0^m,04$
0, 025	7, 44	−0, 0018	7, 32	7, 30	7, 35	−0, 0016	−0, 04	−0, 06	−0, 01
0, 050	7, 46	−0, 0010	7, 35	7, 32	7, 37	−0, 0008	−0, 01	−0, 04	+0, 01
0, 075	7, 48	−0, 0004	7, 37	7, 35	7, 38	−0, 0002	+0, 01	−0, 01	+0, 02
0, 100	7, 50	+0, 0004	7, 38	7, 37	7, 39	+0, 0006	+0, 02	+0, 01	+0, 03
0, 125	7, 51	+0, 0014	7, 39	7, 38	7, 41	+0, 0016	+0, 03	+0, 02	+0, 05
0, 150	7, 51	+0, 0021	7, 41	7, 39	7, 41	+0, 0023	+0, 05	+0, 03	+0, 05
0, 175	7, 50	+0, 0034	7, 41	7, 41	7, 41	+0, 0036	+0, 05	+0, 05	+0, 05
0, 200	7, 49	+0, 0046	7, 41	7, 41	7, 41	+0, 0048	+0, 05	+0, 05	+0, 05
0, 225	7, 48	+0, 0056	7, 41	7, 41	7, 43	+0, 0058	+0, 05	+0, 05	+0, 07
0, 250	7, 45	+0, 0064	7, 43	7, 41	7, 44	+0, 0066	+0, 07	+0, 05	+0, 08
0, 275	7, 43	+0, 0066	7, 44	7, 43	7, 46	+0, 0068	+0, 08	+0, 07	+0, 10
0, 300	7, 41	+0, 0067	7, 46	7, 44	7, 48	+0, 0069	+0, 10	+0, 08	+0, 10
0, 325	7, 39	+0, 0065	7, 48	7, 46	7, 50	+0, 0067	+0, 12	+0, 10	+0, 14
0, 350	7, 37	+0, 0062	7, 50	7, 48	7, 51	+0, 0064	+0, 14	+0, 12	+0, 15
0, 375	7, 34	+0, 0058	7, 51	7, 50	7, 51	+0, 0060	+0, 15	+0, 14	+0, 15
0, 400	7, 32	+0, 0050	7, 51	7, 51	7, 50	+0, 0052	+0, 15	+0, 15	+0, 14
0, 425	7, 30	+0, 0042	7, 50	7, 51	7, 49	+0, 0044	+0, 14	+0, 15	+0, 13
0, 450	7, 29	+0, 0035	7, 49	7, 50	7, 48	+0, 0037	+0, 13	+0, 14	+0, 12
0, 475	7, 27	+0, 0026	7, 48	7, 49	7, 45	+0, 0028	+0, 12	+0, 13	+0, 09
0, 500	7, 25	+0, 0016	7, 45	7, 48	7, 43	+0, 0018	+0, 09	+0, 12	+0, 07
0, 525	7, 23	+0, 0005	7, 43	7, 45	7, 41	+0, 0007	+0, 07	+0, 09	+0, 05
0, 550	7, 21	+0, 0004	7, 41	7, 43	7, 39	−0, 0002	+0, 05	+0, 07	+0, 03
0, 575	7, 20	−0, 0014	7, 39	7, 41	7, 37	−0, 0012	+0, 03	+0, 05	+0, 01
0, 600	7, 19	−0, 0022	7, 37	7, 39	7, 34	−0, 0020	+0, 01	+0, 03	−0, 02
0, 625	7, 18	−0, 0030	7, 34	7, 37	7, 32	−0, 0028	−0, 02	+0, 01	−0, 04
0, 650	7, 20	−0, 0035	7, 32	7, 34	7, 30	−0, 0033	−0, 04	−0, 02	−0, 06
0, 675	7, 23	−0, 0040	7, 30	7, 32	7, 29	−0, 0038	−0, 06	−0, 04	−0, 07
0, 700	7, 24	−0, 0044	7, 29	7, 30	7, 27	−0, 0042	−0, 07	−0, 06	−0, 09
0, 725	7, 27	−0, 0049	7, 27	7, 29	7, 25	−0, 0047	−0, 09	−0, 07	−0, 11
0, 750	7, 30	−0, 0051	7, 25	7, 27	7, 23	−0, 0049	−0, 11	−0, 09	−0, 13
0, 775	7, 32	−0, 0053	7, 23	7, 25	7, 21	−0, 0051	−0, 13	−0, 11	−0, 15
0, 800	7, 35	−0, 0055	7, 21	7, 23	7, 20	−0, 0053	−0, 15	−0, 13	−0, 16
0, 825	7, 37	−0, 0056	7, 20	7, 21	7, 19	−0, 0054	−0, 16	−0, 15	−0, 17
0, 850	7, 38	−0, 0056	7, 19	7, 20	7, 18	−0, 0054	−0, 17	−0, 16	−0, 18
0, 875	7, 39	−0, 0056	7, 18	7, 19	7, 20	−0, 0054	−0, 18	−0, 17	−0, 16
0, 900	7, 41	−0, 0054	7, 20	7, 18	7, 23	−0, 0052	−0, 16	−0, 18	−0, 13
0, 925	7, 41	−0, 0051	7, 23	7, 20	7, 24	−0, 0049	−0, 13	−0, 16	−0, 12
0, 950	7, 41	−0, 0050	7, 24	7, 23	7, 27	−0, 0048	−0, 12	−0, 13	−0, 09
0, 975	7, 41	−0, 0044	7, 27	7, 24	7, 30	−0, 0042	−0, 09	−0, 12	−0, 06
	7, 36	−0, 0002*							

* = Aici ultima linie conţine valorile medii ale coloanelor respective.

Scanned with OKEN Scanner

Tabela 3.2 (continuare)

$(y-\bar{y})^2$	$(x-\bar{x})_0^2$	$(y-\bar{y})(x-\bar{x})_0$	$(y-\bar{y})(x-\bar{x})_-$	$(y-\bar{y})(x-\bar{x})_+$
0,00000484	0,0036	+0,000132	+0,000198	+0,000088
256	16	+ 24	+ 96	+ 16
064	01	+ 08	+ 32	− 08
004	01	− 02	+ 02	− 04
036	04	+ 12	+ 06	+ 18
256	09	+ 48	+ 32	+ 80
529	25	+ 115	+ 69	+ 115
1296	25	+ 180	+ 180	+ 180
2304	25	+ 240	+ 240	+ 240
3364	25	+ 290	+ 290	+ 406
4356	49	+ 462	+ 330	+ 528
4624	64	+ 544	+ 476	+ 680
4761	100	+ 690	+ 552	+ 828
4489	144	+ 804	+ 670	+ 938
4096	196	+ 896	+ 768	+ 960
3600	225	+ 900	+ 840	+ 900
2704	225	+ 780	+ 780	+ 728
1936	196	+ 616	+ 660	+ 572
1369	169	+ 481	+ 518	+ 444
784	144	+ 336	+ 364	+ 252
324	81	+ 162	+ 216	+ 126
049	49	+ 49	+ 263	+ 35
004	25	− 10	− 14	− 06
144	09	− 36	− 60	− 12
400	01	− 20	− 60	+ 40
784	04	+ 56	− 28	+ 112
1089	16	+ 132	+ 66	+ 198
1444	36	+ 228	+ 152	+ 266
1764	49	+ 294	+ 252	+ 378
2209	81	+ 423	+ 329	+ 517
2401	121	+ 539	+ 441	+ 637
2601	169	+ 663	+ 561	+ 765
2809	225	+ 795	+ 689	+ 848
2916	256	+ 864	+ 810	+ 918
2916	289	+ 918	+ 864	+ 972
2916	324	+ 972	+ 918	+ 864
2704	256	+ 832	+ 936	+ 676
2401	169	+ 637	+ 784	+ 588
2304	144	+ 576	+ 624	+ 432
1764	81	+ 378	+ 504	+ 252
0,00075255	0,4064	+0,016008	+0,015150	+0,016567*

* = Aici ultima linie conține sumele coloanelor respective.

Scanned with OKEN Scanner

Aplicînd formula (3.41), pentru cele trei serii de date, rezultă

$$r_0 = \frac{0{,}016008}{\sqrt{0{,}00075255 \times 0{,}4064}} = \frac{0{,}016008}{0{,}017488} = 0{,}9154.$$

$$r_- = \frac{0{,}015150}{0{,}017488} = 0{,}8663$$

$$r_+ = \frac{0{,}016567}{0{,}017488} = 0{,}9473$$

Valorile numerice calculate aici le înlocuim în formula (3.66) și avem

$$\varphi = -0^P_{,}250 - 0^P_{,}0125 \frac{0{,}9473 - 0{,}8663}{0{,}9473 + 0{,}8663 - 0{,}9154 \times 2}$$

adică, în cazul prezent, faza de corelație este

$$\varphi = -0^P_{,}191$$

ii) În cazul cînd x_k, y_k sînt exprimate în același sistem de unități, determinarea fazei de corelație se poate efectua și fără a calcula în mod efectiv mărimea coeficientului de corelație. În acest scop, de pe graficul celor două funcții $x_k = f(t_k)$ și $y = g(t_k)$ ne putem da seama cam cu cît ar trebui să deplasăm abscisele punctelor funcției $y_k = g(t_k)$ pentru ca extremele celor două curbe să vină la aceeași abscisă. În cazul cînd curbele nu au extreme bine puse în evidență, ne putem pune aceeași problemă cu privire la translația uneia dintre cele două curbe pînă cînd curbele respective devin paralele.

Să considerăm că în $x_k = f(t_k)$ menținem fixe abscisele t_k, iar în $y_k = g(t_k)$, prin evaluarea de pe grafic, aplicăm corecția Δt_0, adică să avem $y_k = g(t_k + \Delta t_0)$.

Am amintit mai înainte că în cercetarea seriilor dinamice, problemele se simplifică mult dacă elementele corespunzătoare sînt echidistante în timp. De aceea, și în acest caz vom considera că Δt_0 este egal cu un număr întreg de pași δt.

Să reprezentăm acum datele sub forma unui tabel în care prima coloană conține momentele t_k, în a doua sînt valorile lui $x_k = f(t_k)$, în a treia coloană se află mărimile funcției $y_k = g(t_k)$, iar în coloana a patra se găsesc mărimile $y_k = g(t_k + N \cdot \delta t)$.

Să calculăm acum diferențele valorilor $x_k - y_k$ și să notăm

$$\left.\begin{aligned} S_0 &= \Sigma(x_k - y_k)^2 \text{ pentru } y_k = g(t_k + N \cdot \delta t) \\ S_- &= \Sigma(x_k - y_k)^2 \text{ pentru } y_k = g[t_k + (N-1) \cdot \delta t] \\ S_+ &= \Sigma(x_k - y_k)^2 \text{ pentru } y_k = g[t_k + (N+1) \cdot \delta t] \end{aligned}\right\} \quad (3.67)$$

Scanned with OKEN Scanner

În cazul cînd $\Delta t_0 (= N \cdot \delta t)$ este destul de apropiat de mărimea fazei de corelație, putem scrie

$$S = A(\Delta t)^2 + B(\Delta t) + C \tag{3.68}$$

unde punem condiția ca S să fie minim pentru $(\Delta t)_{min}$. Prin urmare,

$$\varphi = (\Delta t)_{min} = -\frac{B}{2A} \tag{3.69}$$

Pentru determinarea coeficienților A, B și C din relațiile (3.68) și (3.69), procedăm exact ca în cazul sistemului (3.63) cu singura deosebire că aici vom avea S în loc de r. Așadar, relațiile (3.64) se vor scrie:

$$\left.\begin{aligned} A = \frac{S_+ + S_- - 2S_0}{2(\delta t)^2}, \qquad B = \frac{S_+ - S_- - 2(S_+ - S_- + 2S_0)N}{2(\delta t)} \\ C = S_0 - A(\Delta t_0)^2 - B(\Delta t_0) \end{aligned}\right\} \tag{3.70}$$

În aceste condiții, pentru faza de corelație, avem

$$\varphi = \frac{2(\delta t)^2}{2(S_+ + S_- - 2S_0)} \cdot \frac{S_+ - S_- - 2(S_+ + S_- - 2S_0)N}{2(\delta t)}$$

sau, ținînd seama de notația $\Delta t_0 = N \cdot \delta t$,

$$\varphi = \Delta t_0 - \frac{\delta t}{2} \cdot \frac{S_+ - S_-}{S_+ + S_- - 2S_0} \tag{3.71}$$

Aplicație practică

Pentru exemplificarea modului de determinare a fazei de corelație în cazul cînd mărimile x_k, y_k sînt exprimate în același sistem de unități, ne folosim de observațiile fotoelectrice ale stelei *VZ Cancri*, efectuate de H. Spinrad (1960). În modul acesta vom nota $x = B - V$, indicele fotometric definit de diferența de magnitudini în domeniul albastru (B) și cel galben (V), pe cînd $y = U - B$ este indicele fotometric definit de domeniile ultraviolet (U) și respectiv albastru (B). Astfel, din datele observaționale care sînt reprezentate grafic pe figura 3.5, am extras mărimile (echidistante în timp) necesare și care sînt date numeric în tabela 3.3. Aici, în primele trei coloane se găsesc momentele observațiilor, t_k, exprimate în zile iuliene (vom reveni asupra acestei noțiuni) și cei doi indici de culoare: $x_k = (B - V)_k$ și $y_k = (U - B)_k$.

De pe graficul din figura 3.5 se vede că maximele celor două curbe s-ar găsi aproximativ la aceeași abscisă dacă pentru abscisele funcției $y_k = g(t_k)$ aplicăm o corecție $\Delta t_0 = +0^d{,}025$, unde d înseamnă ziua iuliană care a fost aleasă ca unitate de timp.

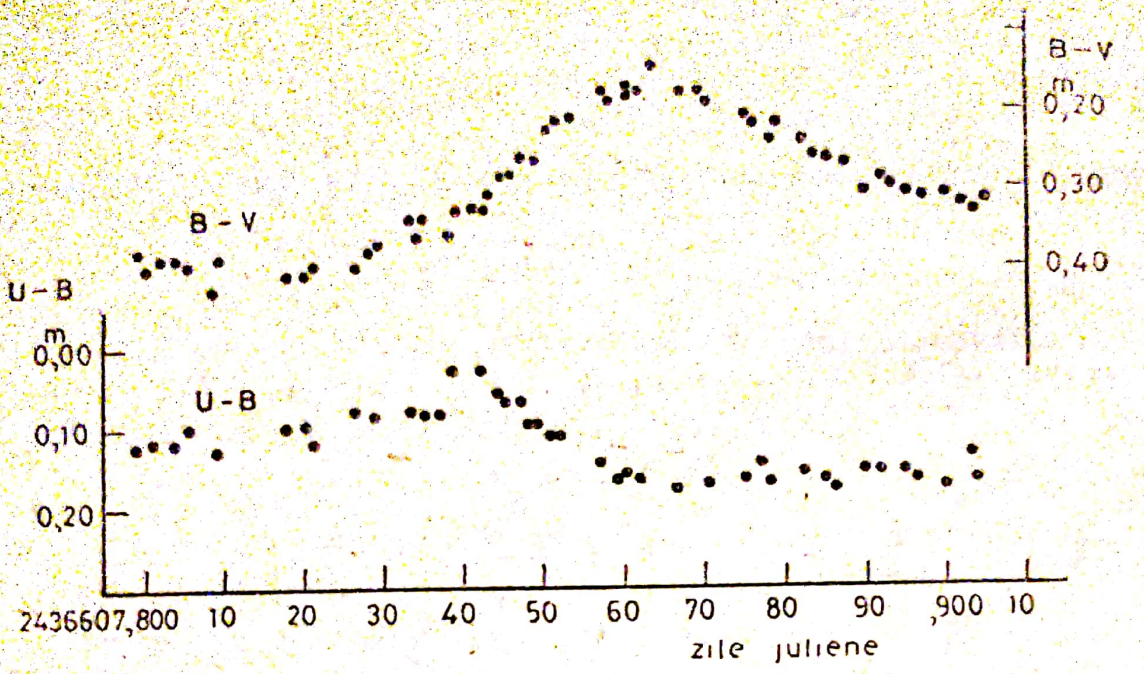


Fig. 3.5. **Determinarea fazei de corelaţie cînd x_k şi y_k sînt exprimate în acelaşi sistem de unităţi.**

Tabela 3.3

t	x	y	y_0	y_+	y_-	$x-y_0$	$x-y_+$	$x-y_-$	$(x-y_0)^2$	$(x-y_+)^2$	$(x-y_-)^2$
2436000+											
607,800	0,39	0,12									
,805	,39	,11									
,810	,40	,11									
,815	,40	,11									
,820	,40	,10									
,825	,40	,10	12	12	11	28	28	29	784	784	841
,830	,37	,09	11	12	11	26	25	26	676	625	676
,835	,35	,07	11	11	11	24	24	24	576	576	576
,840	,33	,03	11	11	10	22	22	23	484	484	529
,845	,29	,06	10	11	10	19	18	19	361	324	361
,850	,23	,11	10	10	09	13	13	14	169	169	196
,855	,20	,14	09	10	07	11	10	13	121	100	169
,860	,18	,16	07	09	03	11	9	15	121	81	225
,865	,16	,18	03	07	06	13	9	10	169	81	100
,870	,19	,18	06	03	11	13	16	8	169	256	64
,875	,21	,17	11	06	14	10	15	7	100	225	49
,880	,24	,16	14	11	16	10	13	8	100	169	64
,885	,26	,17	16	14	18	10	12	8	100	144	64
,890	,29	,16	18	16	18	11	13	11	121	169	121
,895	,30	,16	18	18	17	12	12	13	144	144	169
,900	,31	,16	17	18	16	14	13	15	196	169	225
,905	,32	,16	16	17	17	16	15	15	256	225	225
									4647	4725	4654

Prin urmare mărimea y de la $t_k = ,800$ se deplasează la $t_k = ,825$ iar y de la $t_k = ,860$ vine la $t_k = ,885$. În modul acesta am obținut coloana a 4-a în care avem mărimile y_0. Aici trebuie să specificăm că, deoarece nu am dispus de un ciclu întreg observat, au rămas unele valori neutilizate, iar în partea de sus a coloanei y_+ nemaiavînd mărimi suficiente, am repetat mărimea 12. De asemenea, trebuie să mai specificăm și faptul că, începînd de la coloana y_0, toate mărimile corespunzătoare le-am scris ca și cum ar fi numere întregi, adică ca și cum am fi înmulțit cu 100 peste tot. Această operație nu influențează rezultatele deoarece acest factor apare atît la numărătorul formulei (3.71) cît și la numitorul respectiv.

Pentru obținerea coloanei a 5-a, la abscisele coloanei a 4-a am aplicat corecția $\delta t = 0,005$, ceea ce înseamnă că la abscisele din coloana a doua aplicăm corecția $\Delta t_+ = 0^d,030$. Prin urmare, y de la abscisa $t_k = ,800$ se va deplasa la $t_k = ,830$, iar y de la abscisa $t_k = ,860$ va trece la abscisa $t_k = ,980$ etc. Avem în acest caz mărimile y_+.

Coloana a 6-a se obține în mod analog, la abscisele din coloana y_0 vom aplica corecția $\delta t = -0^d,005$ ceea ce înseamnă că la abscisele mărimilor din coloana a doua aplicăm corecția $\Delta t_- = 0^d,020$. Prin urmare, y de la $t_k = ,805$ se va deplasa la $t_k = ,825$, iar y de la $t_k = ,860$ se va deplasa la $t_k = ,880$ etc. Avem în acest caz mărimile y_-. Conținutul următoarelor șase coloane este specificat în capul fiecărei coloane, iar sumele ultimelor trei coloane ne dau mărimile S_0, S_+ și respectiv S_-. Pe acestea din urmă le înlocuim în formula (3.71) și avem

$$\varphi = +0^d,025 - 0^d,0025 \frac{4725 - 4654}{4725 + 4654 - 2 \times 4647}$$

adică

$$\varphi = +0,023 \text{ zile iuliene.}$$

Observație

Din determinarea fazei de corelație între cei doi indici de culoare pot să apară probleme foarte interesante. Astfel, de exemplu, s-ar putea pune în evidență anumite cauze fizice care să exprime faptul că în diferite domenii spectrale, maximele de strălucire se produc la momente diferite.

b. *Autocorelația — Corelația de serie*

În cadrul seriilor dinamice oscilatorii cum sînt, de exemplu, funcțiile periodice, este deosebit de importantă noțiunea de *autocorelație* care, de fapt, reprezintă corelația dintre termenii unei aceleiași serii. Prin urmare, autocorelația sau corelația internă — core-

Scanned with OKEN Scanner

lația de serie — este o corelație între valorile luate de una și aceeași variabilă la diferite momente separate între ele printr-un anumit interval de timp. Astfel, dacă ne referim, de exemplu, la formula (3.41b), vom numi *corelație de serie de primul ordin* cazul în care coeficientul de corelație va fi definit prin formula

$$r = \frac{\sum_{k=1}^{n} x_k x_{k-1}}{\sum_{k=1}^{n} x_k^2} \tag{3.72}$$

unde, la numitorul fracției respective, am putut considera $\Sigma x_k^2 = \Sigma x_{k-1}^2$. În mod analog, pentru autocorelația de ordinul j vom avea

$$r = \frac{\sum_{k=1}^{n} x_k x_{k-j}}{\sum_{k=1}^{n} x_k^2} \tag{3.72a}$$

unde r este acum un coeficient de corelație de serie ciclică. Acest raționament rămîne valabil și pentru cazul în care mediile $\bar{x}$ și $\bar{y}$ nu sînt nule, pentru autocorelație trebuind să calculăm valorile respective pentru intervalele corespunzătoare de timp. În cazul funcțiilor periodice, valoarea parametrului j va indica ordinul de mărime al perioadei seriei examinate.

Să considerăm o serie oscilatorie al cărei grafic reprezintă o curbă „cvasi-periodică" (vezi figura 3.6). În acest caz va exista un

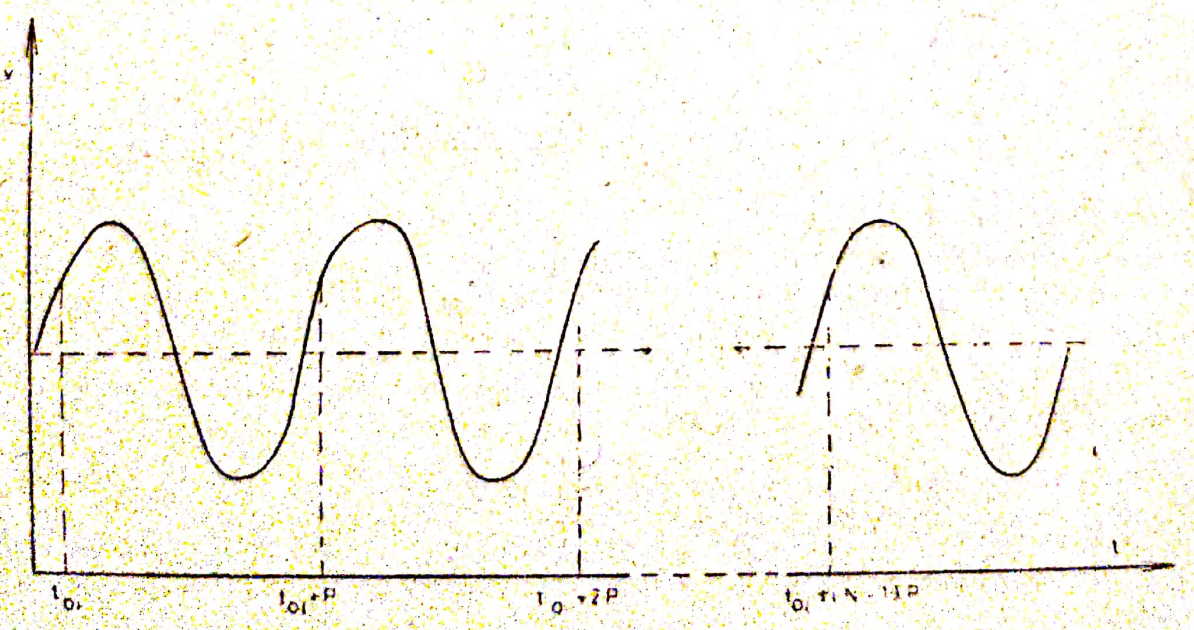


Fig. 3.6. Definirea autocorelației în cadrul unei funcții periodice

Scanned with OKEN Scanner

interval de timp egal cu perioada P după care „se repetă" curba respectivă și, așa cum am specificat și mai înainte, punctele respective fiind situate la abscise echidistante, se poate presupune că avem același număr de puncte în fiecare ciclu observat.

Fie n numărul de puncte existente într-un ciclu ($i = \overline{1, n}$), și N numărul de cicluri.

În aceste condiții, pornind de la expresia coeficientului de corelație (vezi formula 3.41), comparînd primul ciclu (y_{oi}) cu toate celelalte cicluri (y_{oi+kP}, $k = \overline{0, N-1}$), putem scrie

$$R = \frac{\sum_{i=1}^{n} \sum_{k=1}^{N-1} (y_{oi} - \bar{y})(y_{oi+kP} - \bar{y})}{\sqrt{(N-1) \sum_{i=1}^{n} (y_{oi} - \bar{y})^2 \sum_{i=1}^{n} \sum_{k=1}^{N-1} (y_{oi+kP} - \bar{y})^2}}$$

sau

$$R = \frac{\sum_{i=1}^{n} (y_{oi} - \bar{y}) \sum_{k=1}^{N-1} (y_{oi+kP} - \bar{y})}{\sqrt{(N-1) \sum_{i=1}^{n} (y_{oi} - \bar{y})^2 \sum_{i=1}^{n} \sum_{k=1}^{N-1} (y_{oi+kP} - \bar{y})^2}} \qquad (3.72b)$$

unde R este coeficientul de corelație de serie ciclică. Factorul $N - 1$ apare la numitorul de mai sus deoarece în funcțiile respective se utilizează mărimea medie a sumelor

$$\sum_{k=1}^{N-1} (y_{oi+kP} - \bar{y}) \quad \text{și} \quad \sum_{k=1}^{N-1} (y_{oi+kP} - \bar{y})^2$$

În cazul cînd o curbă este strict periodică și am dispus de valoarea exactă a perioadei P, avem

$$y_{oi} = y_{oi+kP}$$

ceea ce înseamnă că din (3.72b) rezultă

$$R = \frac{\sum_{i=1}^{n} (y_{oi} - \bar{y})(N-1)(y_{oi} - \bar{y})}{\sqrt{(N-1) \sum_{i=1}^{n} (y_{oi} - \bar{y})^2 (N-1) \sum_{i=1}^{n} (y_{oi} - \bar{y})^2}} = \pm 1 \qquad (3.73)$$

În cazul cînd urmează să cercetăm autocorelația dintre două cicluri oarecare, avem

$$R = \frac{\sum_{i=1}^{n} (y_{oi} - \bar{y})(y_{oi+kP} - \bar{y})}{\sqrt{\sum_{i=1}^{n} (y_{oi} - \bar{y})^2 \sum_{i=1}^{n} (y_{oi+kP} - \bar{y})^2}} \qquad (3.73a)$$

Scanned with OKEN Scanner

unde $k = \overline{0 ; N-1}$ este un număr constant și indică numărul de ordine al celui de al doilea ciclu utilizat în comparație cu primul ciclu.

c. *Utilizarea autocorelației pentru determinarea perioadei unei funcții empirice periodice*

În cadrul formulelor de mai sus am presupus cunoscută perioada P a funcției respective. Dar dacă acest parametru este necunoscut, el se poate determina printr-o metodă în care să folosim un procedeu de aproximații succesive. În acest scop vom urma un raționament identic cu acela prezentat în paragraful 3.10a : vom atribui trei valori numerice care să difere prin pasul ΔP = constant și destul de mic. În acest sens vom calcula mărimile : R_0 pentru $\Delta P = 0$, R_+ pentru $\Delta P > 0$ și R_- pentru $\Delta P < 0$. Apoi, făcînd același raționament ca și la stabilirea formulelor (3.66) și (3.71) vom scrie

$$P = \tilde{P} + \delta P$$

unde $\tilde{P}$ este o mărime aproximativă pe care am evaluat-o de pe graficul funcției de cercetat, iar corecția acesteia, δP, este definită de

$$\delta P = -\frac{\Delta P}{2} \cdot \frac{R_+ - R_-}{R_+ - R_- - 2R_0} \qquad (3.74)$$

Aplicație numerică

Pentru a ilustra modul de utilizare practică a considerațiilor prezentate mai sus, dăm în tabela 3.4a datele numerice care definesc o curbă empirică ; aici x_i reprezintă timpul exprimat în zile și fracțiuni de zi, iar y_i sînt ordonatele corespunzătoare. În cadrul tabelelor 3.4b, 3.4c și 3.4d sînt determinate mărimile R_0, R_+ și respectiv R_-. Aici se vede că din a doua coloană a tabelei 3.4a, cu excepția primei și ultimei mărimi, au fost utilizate toate ordonatele care au fost comparate cu ordonatele din ciclurile următoare.

Cînd am utilizat $\Delta P = 0$ au rămas neutilizate primele și ultimele ordonate din toate ciclurile ; cînd am utilizat $\Delta P > 0$ din ultimele trei cicluri au rămas primele două ordonate neutilizate, iar pentru $\Delta P < 0$ au rămas ultimele două ordonate neutilizate. Aceste situații se datoresc faptului că primul ciclu este comparat cu celelalte cicluri în funcție de corecția ΔP utilizată. Valoarea adoptată a perioadei aproximative este $P = 0^d,650$, timpul în abscisă fiind exprimat în zile (d), iar $\Delta P = 0^d,01$.

Coloanele tabelelor au semnificația specificată la capul corespunzător, iar formulele scrise sub fiecare tabelă arată operațiile care au fost executate pentru calcularea celor trei mărimi : R_0, R_+, și R_-. Din compararea rezultatelor astfel obținute se vede că între primul ciclu și ciclurile următoare există o corelație bună, dar pentru

Scanned with OKEN Scanner

$\Delta P = 0^d,00$ se obține cea mai mare valoare a coeficientului de corelație $R_0 = 0{,}9982$. De altfel, utilizînd expresia (3.74), avem

$$\delta P = -0{,}005 \frac{0{,}9317 - 0{,}9397}{0{,}9317 + 0{,}9397 - 2(0{,}9982)} = -0^d,00032$$

iar din (3.75) rezultă

$$P = 0^d,650 - 0^d,00032 = 0^d,64968.$$

(aici litera d ne arată că timpul este exprimat în zile).

Tabela 3.4a (Curba empirică periodică)

x	y	x	y	x	y	x	y
210,210	0,00	210,860	0,01	211,510	0,00	212,160	0,01
,220	,05	,870	,04	,520	,05	,170	,06
,230	,10	,880	,09	,530	,09	,180	,10
,240	,15	,890	,15	,540	,14	,190	,15
,250	,20	,900	,18	,550	,19	,200	,19
,260	,25	,910	,26	,560	,25	,210	,26
,270	,30	,920	,31	,570	,30	,220	,31
,280	,35	,930	,35	,580	,36	,230	,36
,290	,40	,940	,42	,590	,41	,240	,40
,300	,45	,950	,46	,600	,45	,250	,46
210,310	0,50	210,960	0,50	211,610	0,49	221,260	0,52

Tabela 3.4b (Calculul lui R_0)

y_0	y_1	y_2	y_3	$y_0-\bar{y}$	$y_1-\bar{y}$	$y_2-\bar{y}$	$y_3-\bar{y}$	$\sum_{k=1}^{N-1}(y_{oi+kP}-\bar{y})$	$\sum_{k=1}^{N-1}(y_{oi+kP}-\bar{y})^2$
0,05	0,04	0,05	0,06	−0,20	−0,21	−0,20	−0,19	−0,60	0,1202
,10	,09	,09	,10	− ,15	− ,16	− ,16	− ,15	− ,47	,0737
,15	,15	,14	,15	− ,10	− ,10	− ,11	− ,10	− ,31	,0321
,20	,18	,19	,19	− ,05	− ,07	− ,06	− ,06	− ,19	,0121
,25	,26	,25	,26	,00	+ ,01	,00	+ ,01	+ ,02	,0002
,30	,31	,30	,31	+ ,05	+ ,06	+ ,05	+ ,06	+ ,17	,0097
,35	,35	,36	,36	+ ,10	+ ,10	+ ,11	+ ,11	+ ,32	,0342
,40	,42	,41	,40	+ ,15	+ ,17	+ ,16	+ ,15	+ ,48	,0770
0,45	0,46	0,45	0,46	+0,20	+0,21	+0,20	+0,21	+0,62	0,1828
					N = 4				0,4874

$$\sum_{i=1}^{n}\left[(y_{oi}-\bar{y})\sum_{k=1}^{N-1}(y_{oi+kP}-\bar{y})\right] = 0{,}4675\,; \qquad \sum_{i=1}^{n}\sum_{k=1}^{N-1}(y_{io+kP}-\bar{y})^2 = 0{,}4874$$

$$\sum_{i=1}^{n}(y_{oi}-\bar{y})^2 = 0{,}1500, \qquad \bar{y} = \frac{9{,}04}{36} = 0{,}25$$

$$R_0 = \frac{0{,}4675}{\sqrt{3(0{,}1500)(0{,}4874)}} = 0{,}9982$$

Scanned with OKEN Scanner

Tabela 3.4c (Calculul lui R_+)

y_0	y_1	y_2	y_3	$y_0-\bar{y}$	$y_1-\bar{y}$	$y_2-\bar{y}$	$y_3-\bar{y}$	$\sum_{k=1}^{N-1}(y_{oi+kP}-\bar{y})$	$\sum_{k=1}^{N-1}(y_{oi+kP}-\bar{y})^2$
0,05	0,09	0,09	0,10	−0,24	−0,20	−0,20	−0,19	−0,59	0,1161
,10	,15	,14	,15	− ,19	− ,14	− ,15	− ,14	− ,43	,0617
,15	,18	,19	,19	− ,14	− ,11	− ,10	− ,10	− ,31	,0321
,20	,26	,25	,26	− ,09	− ,03	− ,04	− ,03	− ,10	,0034
,25	,31	,30	,31	− ,04	+ ,02	+ ,01	+ ,02	+ ,05	,0009
,30	,35	,36	,36	+ ,01	+ ,06	+ ,07	+ ,07	+ ,20	,0134
,35	,42	,41	,40	+ ,06	+ ,13	+ ,12	+ ,11	+ ,36	,0434
,40	,46	,45	,46	+ ,11	+ ,17	+ ,16	+ ,17	+ ,50	,0834
0,45	0,50	0,49	0,52	+0,16	+0,21	+0,20	+0,23	+0,64	0,1370
				N=4					0,4914

$$\sum_{i=1}^{n}\left[(y_{oi}-\bar{y})\sum_{k=1}^{N-1}(y_{oi+kP}-\bar{y})\right]=0{,}4587\,;\qquad \sum_{i=1}^{n}\sum_{k=1}^{N-1}(y_{oi+kP}-\bar{y})^2=0{,}4914$$

$$\sum_{i=1}^{n}(y_{oi}-y)^2=0{,}1644,\qquad \bar{y}=\frac{10{,}40}{36}=0{,}29$$

$$R_+=\frac{0{,}4587}{\sqrt{3(0{,}1644)(0{,}4914)}}=0{,}9317$$

Tabela 3.4d (Calculul lui R_-)

y_0	y_1	y_2	y_3	$y_0-\bar{y}$	$y_1-\bar{y}$	$y_2-\bar{y}$	$y_3-\bar{y}$	$\sum_{k=1}^{N-1}(y_{oi+kP}-\bar{y})$	$\sum_{k=1}^{N-1}(y_{oi+kP}-\bar{y})^2$
0,05	0,01	0,00	0,01	−0,16	−0,20	−0,21	−0,20	−0,61	0,1281
,10	,04	,05	,06	− ,11	− ,17	− ,16	− ,15	− ,48	,0770
,15	,09	,09	,10	− ,06	− ,12	− ,12	− ,11	− ,35	,0409
,20	,15	,14	,15	− ,01	− ,06	− ,07	− ,06	− ,19	,0121
,25	,18	,19	,19	− ,04	− ,03	− ,02	− ,02	− ,07	,0017
,30	,26	,25	,26	+ ,09	+ ,05	+ ,04	+ ,05	+ ,14	,0066
,35	,31	,30	,31	+ ,14	+ ,10	+ ,09	+ ,10	+ ,29	,0281
,40	,35	,36	,36	+ ,19	+ ,14	+ ,15	+ ,15	+ ,44	,0646
0,45	0,42	0,41	0,40	+0,24	+0,21	+0,20	+0,19	+0,60	0,1202
				N = 4					0,4793

$$\sum_{i=1}^{n}\left[(y_{oi}-\bar{y})\sum_{k=1}^{N-1}(y_{oi+kp}-y)\right]=0{,}4569\,;\qquad \sum_{i=1}^{n}\sum_{k=1}^{N-1}(y_{oi+kP}-\bar{y})^2=0{,}4793$$

$$\sum_{i=1}^{n}(y_{oi}-\bar{y})^2=0{,}1644,\qquad \bar{y}=\frac{7{,}69}{36}=0{,}21$$

$$R_-=\frac{0{,}4569}{\sqrt{3(0{,}1644)(0{,}4793)}}=0{,}9397$$

Scanned with OKEN Scanner

3.10. Corelația neliniară. Raportul de corelație

Am văzut că, în problemele de corelație, dreapta de regresie era caracteristică pentru o dependență funcțională liniară. Dar, așa cum am mai amintit, lipsa unei dependențe liniare ($r = 0$) nu înseamnă că între mărimile x_k și y_k nu există nici un fel de corelație.

În activitatea practică se întîlnesc foarte multe cazuri în care corelația dintre anumiți parametri fizici se poate exprima printr-o curbă oarecare. Prin urmare, se pune problema de a stabili astfel de corelații și de a determina abaterea lor de la liniaritate. În acest scop vom scrie tabela (X) sub forma următoare (vezi tabela X'),

Tabela (X')

$x^{(1)}$	y_{11},	y_{12},	y_{13},	...	y_{1m}	$\bar{y}_{x1}$	$\sigma(\bar{y}_{x1})$
$x^{(2)}$	y_{21},	y_{22},	y_{23},	...	y_{2m}	$\bar{y}_{x2}$	$\sigma(\bar{y}_{x2})$
$x^{(3)}$	y_{31},	y_{32},	y_{33},	...	y_{3m}	$\bar{y}_{x3}$	$\sigma(\bar{y}_{x3})$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
$x^{(l)}$	y_{l1},	y_{l2},	y_{l3},	...	y_{lm}	$\bar{y}_{xl}$	$\sigma(\bar{y}_{xl})$

În ultimele două coloane ale tabelei (X') se găsesc valorile medii ale mărimilor y de pe linia respectivă și deviațiile standard corespunzătoare.

Pentru a calcula valoarea medie $\bar{y}$ și deviația standard a tuturor mărimilor y ne folosim de formulele (3.8) și respectiv (3.9). Apoi, din penultima coloană determinăm media $\bar{y} = \bar{\bar{y}}$ a valorilor medii $\bar{y}_x$ și deviația standard $\sigma(\bar{y}_x)$. Pe baza acestor considerații se introduce noțiunea de raport de corelație prin

$$\eta_{y/x} = \frac{\sqrt{\dfrac{\sum_{i=1}^{l} n_i [(\bar{y}_x)_i - \bar{y}]^2}{\sum_{i=1}^{l} n_i}}}{\sqrt{\dfrac{\sum_{k=1}^{n} n_k (y_k - \bar{y})^2}{\sum_{k=1}^{n} n_k}}} = \frac{\sigma(\bar{y}_x)}{\sigma(y)} \tag{3.75}$$

unde n_l este numărul de valori y care se găsesc pe linia l-a iar $\sum_{k=1}^{n} n_k$ reprezintă numărul total de puncte de pe toate liniile şi toate coloanele. Prin urmare, pentru raportul de corelaţie avem

$$\eta_{y/x} = \sqrt{\frac{\sum_{i=1}^{l} n_i[(\bar{y}_x)_i - \bar{y})]^2}{\sum_{k=1}^{n} n_k (y_k - \bar{y})^2}} \tag{3.76}$$

unde n_k este numărul de diferenţe dintre valorile individuale şi mărimea medie corespunzătoare, adică numărul de diferenţe $y_k - \bar{y}$.

Notaţia $\eta_{y/x}$ înseamnă raportul de corelaţie al lui y ca funcţie de x. În mod analog putem scrie

$$\eta_{x/y} = \frac{\sigma(x_y)}{\sigma(x)} = \frac{1}{\sigma(x)} \sqrt{\frac{1}{n_m} \sum_{j=1}^{m} n_j [(\bar{x}_y)_j - \bar{x})]^2} \tag{3.77}$$

În cele ce urmează ne vom ocupa numai de raportul $\eta_{y/x}$ pe care-l vom nota simplu η.

Din tabela (X') se vede că dacă valorile σ vor fi mici, înseamnă că elementele de pe linia respectivă vor avea aproximativ aceeaşi mărime, dar pot diferi mult de la o linie la alta. Această împrejurare ar putea duce la o corelaţie între mărimile x_i şi mediile corespunzătoare $\bar{y}_{xi}$ care se poate pune în evidenţă prin determinarea raportului de corelaţie η.

Valoarea medie a deviaţiilor σ_{xi} va fi definită de

$$\bar{\sigma}_x^2(y) = \sum_{i=1}^{l} \sigma_{x_i}^2(y) f_i = \sum_{i=1}^{l} \left\{ \sum_{j=1}^{m} (y_{ji} - \bar{y}_{x_i})^2 f_j \right\} f_i \tag{3.78}$$

unde f_i şi f_j sînt frecvenţele corespunzătoare definite de relaţia (3.1).

Pe de altă parte, dacă adunăm toate valorile y se poate obţine media generală $\bar{y}$ cu deviaţia respectivă σ_y. În acest caz, adunînd mai întîi orizontal şi apoi vertical, putem scrie

$$\sigma_y^2 = \sum_{j=1}^{m} \left\{ \sum_{i=1}^{l} (y_{ji} - \bar{y})^2 f_i \right\} f_j \tag{3.79}$$

sau, utilizînd substituţia

$$y_{ji} - \bar{y} = (y_{ji} - \bar{y}_{xi}) + (\bar{y}_{x_i} - \bar{y}) \tag{3.80}$$

putem scrie mai departe

$$\sigma_y^2 = \sum_{j=1}^{m} \left\{ \sum_{i=1}^{l} (y_{ji} - \bar{y}_{xi})^2 f_i \right\} f_j + \sum_{j=1}^{m} \left\{ \sum_{i=1}^{l} (\bar{y}_{xi} - \bar{y})^2 f_i \right\} f_j \tag{3.81}$$

În formula (3.81) am neglijat termenul cu suma dublelor produse deoarece valorile $(\bar{y}_{ji} - \bar{y}_{xi})$ pozitive sînt la fel de probabile ca și cele negative. Această aproximație este cu atît mai valabilă, cu cît este mai mare numărul datelor observaționale. Așadar,

$$\sum_{i=1}^{l} (\bar{y}_{xi} - \bar{y})^2 f_i \sum_{j=1}^{m} f_j = \sum_{i=1}^{l} (\bar{y}_{xi} - \bar{y})^2 f_i = \sigma^2(\bar{y}_x) \tag{3.82}$$

deoarece

$$\sum_{j=1}^{m} f_i = \sum_{j=1}^{m} (n_j/n) = 1, \qquad n = n_0 + n_1 + \ldots$$

Ținînd seamă de relațiile (3.78) și (3.82), expresia (3.81) devine

$$\sigma_y^2 = \bar{\sigma}_x^2(y) + \sigma^2(\bar{y}_x) \tag{3.83}$$

Considerînd raportul de corelație definit de (3.75), avem

$$(1 - \eta^2)\sigma_y^2 = \sigma_y^2 - \sigma^2(\bar{y}_x) \tag{3.84}$$

sau, ținînd seama de (3.83), putem scrie mai departe

$$(1 - \eta^2)\sigma_y^2 = \bar{\sigma}_x^2(y) \tag{3.85}$$

Observații :

a) Dacă $\eta = 0$, rezultă că $\sigma_y^2 = \bar{\sigma}_y^2(y)$, ceea ce înseamnă că nu există o legătură de corelație.

b) Dacă $\eta = 1$, rezultă $\sigma_x^2(y) = 0$, ceea ce înseamnă că trebuie să existe $\bar{\sigma}_x(y) = 0$. Dar aceasta înseamnă că elementele y de pe o linie a tabelei (X') trebuie să fie egale, adică la un anumit x corespunde un anumit y. Aceasta înseamnă că pentru $\eta = 1$, corelația este exprimată printr-o dependență funcțională. Prin urmare, pentru raportul de corelație avem următoarele inegalități

$$0 \leqslant \eta \leqslant 1 \tag{3.86}$$

c) Pentru evaluarea deviației standard a raportului de corelație se poate utiliza formula aproximativă (vezi Romanovschi, 1961, pg. 386)

$$\sigma_\eta = \frac{1 - \eta^2}{\sqrt{n - 1}} \tag{3.87}$$

d) În cazul cînd avem de-a face cu o regresie liniară, raportul de corelație este egal cu coeficientul de corelație. Acest lucru rezultă imediat dacă în prima relație din (3.46) ridicăm la pătrat, înmulțim cu n_k/n și însumăm. Prin urmare

$$\sum_{k=1}^{n} (\bar{y}_x - \bar{y}_k)^2 \frac{n_k}{n} = \sum_{k=1}^{n} r^2 \cdot \frac{\sigma_y^2}{\sigma_x^2} (x - \bar{x}_k)^2 \frac{n_k}{n}.$$

Scanned with OKEN Scanner

Aici ţinem seamă de relaţia (3.5) care defineşte deviaţia standard, astfel încît

$$\sigma^2(\bar{y}_x) = r^2 \frac{\sigma_y^2}{\sigma_x^2} \sigma_x^2 \qquad (*)$$

Pe de altă parte, raportul de corelaţie fiind definit de formula (3.75) putem scrie mai departe

$$\eta^2 = \frac{\sigma^2(\bar{y}_x)}{\sigma_y^2} \Rightarrow \sigma^2(\bar{y}_x) = \eta^2 \sigma_y^2 \qquad (**)$$

Din relaţiile (*) şi (**) rezultă imediat

$$\eta^2 \sigma_y^2 = r^2 \sigma_y^2$$

sau

$$\eta = r$$

Aplicaţie numerică

În scopul aplicării practice a consideraţiilor prezentate mai sus, am utilizat datele observaţionale care au fost obţinute la staţia meteorologică din Cluj-Napoca. Aceste date sînt reproduse în tabela 3.5a unde în prima coloană se găseşte ziua şi luna din cursul anului 1976 cînd au fost efectuate observaţiile respective, în coloana a 2-a se găseşte numărul de ordine al zilei în cursul anului 1976 şi va fi considerat ca abscisă, x, a datelor experimentale, iar în coloana a 3-a se găseşte temperatura medie zilnică, y, exprimată în grade Celsius.

Pentru examinarea corelaţiei liniare dintre temperatura medie zilnică, y, şi zilele corespunzătoare, x, ne-am servit de formula coeficientului de corelaţie.

$$r = \frac{\sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{n} (x_k - \bar{x})^2 \sum_{k=1}^{n} (y_k - \bar{y}^2)}} \qquad (3.41)$$

Deoarece în tabela 3.5a, abscisele sînt echidistante, avem

$$\bar{x} = \frac{1}{n} \sum_{k=}^{n} x_k = \frac{n(n+1)}{2n} = \frac{n+1}{2} = \frac{367}{2} = 183{,}5$$

unde $n = 366$ este numărul de zile pe care le-a avut anul 1976.

Considerînd acum că deviaţiile $x_k - \bar{x}$ sînt reprezentate atît prin numerele pozitive 0,5, 1,5, ..., 182,5 cît şi prin cele negative, putem scrie

$$\sum_{k=1}^{n} (x_k - \bar{x})^2 = 2 \sum_{k=1}^{N} K^2 = 2 \frac{N(N+1)(2N+1)}{6} = \frac{182{,}5 \cdot 183{,}5 \cdot 366}{3}$$

adică

$$\sum_{k=1}^{n} (x_k - \bar{x})^2 = 4085627 \qquad N = 182{,}5$$

Pentru celelalte mărimi care intervin în formula (3.41) utilizăm procedeele obișnuite de calcul și obținem:

$$\bar{y} = \frac{1}{n}\sum_{k=1}^{n} y_k = \frac{1}{366} \cdot 2780{,}9 = 7{,}6$$

$$\sum_{k=1}^{n} (y_k - \bar{y})^2 = 27818; \qquad \sum_{k=1}^{n} (x_k - \bar{x})(y_k - \bar{y}) = 157804.$$

Prin urmare

$$r = \frac{157804}{\sqrt{4085627 \cdot 27818}} = 0{,}468$$

Faptul că r este mai apropiat de zero decît de unitate, înseamnă că nu se poate vorbi despre existența unei corelații liniare între cele două feluri de mărimi x și y.

În vederea cercetării prezenței unei corelații curbilinii, ne servim de formula raportului de corelație

$$\eta = \sqrt{\frac{\sum_{i=1}^{l} n_i[(\bar{y}_x)_i - \bar{y}]^2}{\sum_{k=1}^{n} n_k(y_k - \bar{y})^2}} \tag{3.76}$$

Pentru utilizarea acestei formule, datele din tabela 3.5a au fost restrînse în grupe de cîte 6 ($n_i = 6$) și s-au calculat valorile medii care sînt date în tabela 3.5b.

Deoarece $n_k = 1$, numitorul din formula (3.76) îl avem calculat de mai înainte, pentru numărător avem $n_i = 6$ iar diferențele $(\bar{y}_x)_i - \bar{y}$ se calculează în mod obișnuit. Astfel, efectuînd înlocuirile valorilor numerice, rezultă.

$$\eta = \sqrt{\frac{4190{,}13 \times 6}{27818}} = \sqrt{0{,}9036}$$

adică

$$\eta = 0{,}95$$

Comparînd această valoare a raportului de corelație ($\eta = 0{,}95$) cu mărimea coeficientului de corelație ($r = 0{,}468$), se constată că între temperatura medie zilnică a aerului din Cluj-Napoca și zilele respective ale anului 1976, există o strînsă corelație care însă nu este de formă liniară, dependența funcțională respectivă avînd o formă mult mai complicată.

Scanned with OKEN Scanner

Tabela 3.5a (Datele observaționale)

Data 1976	x	y	Data 1976	x	y	Data 1976	x	y	Data 1976	x	y
1.I.	1	− 3,8	16.II	47	0,5	1.IV	92	7,8	18.V	139	14,0
2	2	0,2	17	48	0,7	2	93	10,8	19	140	14,7
3	3	− 0,2	18	49	0,0	3	94	10,7	20	141	14,4
4	4	0,0	19	50	−0,4	4	95	12,2	21	142	13,5
5	5	− 6,2	20	51	−2,2	5	96	12,6	22	143	13,9
6	6	− 4,4	21	52	−1,3	6	97	13,6	23	144	13,6
7	7	− 5,0	22	53	−1,5	7	98	12,2	24	145	15,1
8	8	−11,7	23	54	−3,1	8	99	10,2	25	146	14,1
9	9	− 5,5	24	55	−2,2	9	100	6,0	26	147	13,4
10	10	− 5,0	25	56	−5,6	10	101	6,8	27	148	16,0
11	11	− 1,0	26	57	−1,7	11	102	6,5	28	149	14,8
12	12	2,8	27	58	1,7	12	103	9,6	29	150	14,1
13	13	4,3	28	59	2,4	13	104	9,5	30	151	15,2
14	14	− 3,4	29.II	60	0,2	14	105	8,3	31.V	152	13,0
15	15	− 0,6				15	106	9,1			
16	16	− 8,7	1.III	61	2,2	16	107	8,4	1.VI	153	10,8
17	17	−12,0	2	62	1,6	17	108	10,6	2	154	11,8
18	18	− 9,2	3	63	−0,4	18	109	10,6	3	155	11,2
19	19	−11,6	4	64	−4,6	19	110	12,3	4	156	10,6
20	20	− 6,6	5	65	−7,2	20	111	12,6	5	157	13,4
21	21	0,0	6	66	−6,3	21	112	11,0	6	158	14,2
22	22	6,7	7	67	−6,4	22	113	9,7	7	159	16,5
23	23	1,2	8	68	−2,2	23	114	12,7	8	160	17,8
24	24	2,4	9	69	1,1	24	115	12,1	9	161	16,9
25	25	− 1,6	10	70	0,6	25	116	11,0	10	162	13,1
26	26	− 2,9	11	71	−2,2	26	117	9,3	11	163	13,1
27	27	− 3,1	12	72	−3,0	27	118	7,4	12	164	15,0
28	28	− 2,6	13	73	−1,4	28	119	6,3	13	165	15,9
29	29	− 4,6	14	74	0,0	29	120	3,2	14	166	15,2
30	30	− 8,2	15	75	0,2	30.IV	121	3,3	15	167	16,4
31.I	31	− 8,8	16	76	3,1				16	168	15,0
			17	77	5,0	1.V	122	5,4	17	169	12,0
1.II	32	− 8,4	18	78	4,8	2	123	9,5	18	170	13,5
2	33	− 5,8	19	79	3,6	3	124	10,6	19	171	16,5
3	34	− 7,2	20	80	1,4	4	125	13,2	20	172	18,1
4	35	− 4,0	21	81	−4,2	5	126	14,5	21	173	19,6
5	36	− 4,4	22	82	−4,5	6	127	14,6	22	174	19,2
6	37	− 6,2	23	83	−1,7	7	128	14,2	23	175	19,1
7	38	−10,8	24	84	3,5	8	129	13,9	24	176	16,2
8	39	−14,5	25	85	4,8	9	130	13,9	25	177	16,6
9	40	−16,7	26	86	3,4	10	131	14,0	26	178	16,4
10	41	−14,2	27	87	6,4	11	132	14,4	27	179	17,8
11	42	−15,6	28	88	3,9	12	133	14,0	28	180	19,9
12	43	−15,1	29	89	4,0	13	134	13,4	29	181	21,6
13	44	− 8,6	30	90	5,8	14	135	13,6	30.VI	182	21,8
14	45	− 0,4	31.III	91	7,3	15	136	9,7			
15.II	46	− 0,9				16	137	11,0	1.VII	183	20,9
						17.V	138	14,0	2.VII	184	20,0

Scanned with OKEN Scanner

Tabela 3.5a (continuare)

Data 1976	x	y	Data 1976	x	y	Data 1976	x	y	Data 1976	x	y
3. VII	185	19,2	18.VIII	231	15,3	3.X	277	16,1	18.XI	323	5,7
4	186	20,2	19	232	16,3	4	278	15,6	19	324	3,4
5	187	20,4	20	233	15,8	5	279	15,9	20	325	2,3
6	188	17,6	21	234	14,6	6	280	13,5	21	326	5,0
7	189	16,9	22	235	14,7	7	281	15,2	22	327	4,2
8	190	18,4	23	236	14,9	8	282	15,7	23	328	1,2
9	191	15,9	24	237	14,5	9	283	13,8	24	329	−4,1
10	192	16,4	25	238	14,2	10	284	13,0	25	330	−2,8
11	193	11,7	26	239	15,6	11	285	12,7	26	331	−5,6
12	194	17,0	27	240	15,9	12	286	12,1	27	332	−3,2
13	195	17,6	28	241	16,4	13	287	13,9	28	333	−1,7
14	196	18,3	29	242	16,4	14	288	13,5	29	334	+0,2
15	197	19,2	30	243	15,7	15	289	13,8	30.XI.	335	−0,6
16	198	17,9	31.VIII	244	16,2	16	290	11,3			
17	199	18,4				17	291	5,1	1.XII	336	+0,1
18	200	18,6	1.IX	245	16,1	18	292	3,8	2	337	4,7
19	201	21,2	2	246	14,6	19	293	4,7	3	338	6,4
20	202	22,9	3	247	16,8	20	294	3,4	4	339	3,0
21	203	23,2	4	248	15,9	21	295	1,3	5	340	7,4
22	204	20,9	5	249	13,4	22	296	1,0	6	341	3,1
23	205	20,1	6	250	10,2	23	297	2,5	7	342	2,0
24	206	21,7	7	251	11,2	24	298	2,7	8	343	6,1
25	207	20,7	8	252	13,4	25	299	2,6	9	344	6,2
26	208	20,8	9	253	12,8	26	300	2,9	10	345	4,9
27	209	19,4	10	254	12,6	27	301	3,1	11	346	−0,4
28	210	17,1	11	255	13,8	28	302	4,9	12	347	−5,6
29	211	15,6	12	256	15,4	29	303	5,8	13	348	−6,5
30	212	16,4	13	257	16,7	30	304	9,1	14	349	−3,9
31.VII	213	17,0	14	258	17,9	31.X.	305	15,9	15	350	−4,1
			15	259	18,0				16	351	−4,7
1.VIII	214	16,6	16	260	17,4	1.XI.	306	9,0	17	352	−0,4
2	215	15,5	17	261	13,0	2	307	6,2	18	353	−0,0
3	216	13,8	18	262	11,0	3	308	9,6	19	354	−1,6
4	217	13,9	19	263	11,8	4	309	7,8	20	355	0,5
5	218	12,9	20	264	11,8	5	310	9,2	21	356	−0,7
6	219	11,8	21	265	11,8	6	311	10,2	22	357	−3,9
7	220	14,4	22	266	11,9	7	312	12,8	23	358	−5,2
8	221	12,2	23	267	12,6	8	313	9,2	24	359	−4,1
9	222	14,2	24	268	10,3	9	314	8,8	25	360	−3,2
10	223	15,7	25	269	10,6	10	315	5,8	26	361	−4,7
11	224	14,8	26	270	10,8	11	316	4,3	27	362	−9,4
12	225	14,7	27	271	9,5	12	317	3,2	28	363	−12,9
13	226	14,8	28	272	10,7	13	318	4,6	29	364	−4,8
14	227	15,1	29	273	14,6	14	319	10,2	30	365	0,0
15	228	15,5	30.IX.	274	16,6	15	320	9,4	31.XII.	366	−4,0
16	229	15,6	1.X.	275	13,7	16	321	7,8			
17.VIII	230	13,8	2.	276	13,4	17.XI.	322	6,4			

Scanned with OKEN Scanner

Tabela 3.5b (Punctele medii)

$\bar{x}$	$\bar{y}_x$	$\bar{x}$	$\bar{y}_x$	$\bar{x}$	$\bar{y}_x$	$\bar{x}$	$\bar{y}_x$
3,5	— 2,4	99,5	9,3	189,5	17,6	279,5	15,3
9,5	— 4,2	105,5	9,3	195,5	17,0	285,5	13,2
15,5	— 4,9	111,5	9,7	201,5	20,9	291,5	7,0
21,5	— 1,3	117,5	8,2	207,5	20,1	297,5	2,2
27,5	— 3,8	123,5	9,4	213,5	15,7	303,5	6,5
33,5	— 5,6	129,5	14,2	219,5	13,2	309,5	9,3
39,5	—13,0	135,5	12,6	225,5	15,1	315,5	6,0
45,5	— 4,0	141,5	14,0	231,5	15,2	321,5	7,2
51,6	— 1,4	147,5	14,6	237,5	15,0	327,5	1,0
57,5	— 0,9	153,5	12,1	243,5	15,9	333,5	—1,8
63,5	— 2,5	159,5	15,3	249,5	13,5	339,5	4,4
69,5	— 2,0	165,5	15,1	255,5	14,9	345,5	0,8
75,5	+ 2,0	171,5	16,5	261,5	13,8	351,8	—2,5
81,5	— 0,3	177,5	17,7	267,5	11,3	357,5	—2,8
87,5	+ 4,7	183,5	20,6	273,5	13,1	363,5	—6,0
93,5	10,2						

3.11. Corelații multiple și parțiale

În paragrafele precedente ne-am ocupat de unele probleme de corelație luînd în considerare prezența a cel mult două serii statistice x_k și y_k din care una putea să aibă influență asupra celeilalte. Dar, în activitatea practică, se cunosc fenomene naturale și procese sociale care sînt determinate de mai multe cauze, fapt care ne face să greșim atunci cînd considerăm serii statistice de numai două caracteristici, dintre care una este considerată ca parametru principal, iar influența posibilă a altor cauze este neglijată. Astfel, de exemplu, în cazul unei structuri sociale, caracteristicile statistice se modifică sub influența unui număr mare de factori dintre care unii ar putea fi esențiali, iar alții mai puțin importanți. De aceea, efectuînd calculele pentru cazurile în care asupra unei caracteristici se consideră influența mai multor factori, vom avea de-a face cu așa numita *corelație multiplă* sau corelația dintre mai multe caracteristici variabile.

Exemplul 1. Dacă se consideră legătura dintre numărul de copii pe familie, venitul mediu al familiei și vîrsta celor căsătoriți, se va constata că numărul copiilor este corelat negativ cu venitul și vîrsta celor căsătoriți. Este un fapt observat că la familiile cu copii mulți venitul este mai mic, (numărul copiilor este mai mare).

Exemplul 2. Se poate constata că există o corelație pozitivă între volumul recoltei dintr-un an și cantitatea de ploaie căzută în perioada respectivă, dar este mai greu de stabilit o corelație între același volum al recoltei și temperatura medie a anului respectiv.

Scanned with OKEN Scanner

Considerarea problemelor de corelație unde avem de-a face cu cel puțin trei variabile sau chiar mai multe se face prin extinderea considerațiilor prezentate cînd există numai două variabile. De aceea, în cazul *regresiei liniare* pentru mai mult de două variabile, se iau în considerare formulele liniare care sînt valabile pentru astfel de situații. Prin urmare, dacă avem n variabile implicate vom scrie

$$X_1 = a_1 + b_2X_2 + b_3X_3 + \ldots + b_nX_n \qquad (3.88)$$

unde X_i este o variabilă pentru care dispunem de o serie de mărimi determinate, x_{ij}, iar $a_1, b_2, b_3, \ldots, b_n$ sînt constante care urmează a fi determinate prin metoda celor mai mici pătrate; vom reveni asupra acestei probleme.

Formula (3.88) reprezintă *ecuația de regresie liniară generalizată* iar coeficienții b_k ne arată corelația existentă între variabilele X_1 și X_k, $k = \overline{2,n}$, de exemplu. Mărimea b_2 indică schimbarea lui X_1 cînd X_2 se modifică cu o unitate iar celelalte variabile ar rămîne nemodificate. Prin urmare, b_2 indică corelația existentă între X_1 și X_2 și se numește *corelație parțială* în care vom avea de-a face cu coeficienții de regresie parțială și corelație parțială.

Problemele corelației parțiale precum și ale celei multiple le vom expune în paragraful prezent. În acest scop va trebui ca a_1 și b_i din (3.88) să fie determinați prin metoda celor mai mici pătrate, iar erorile de determinare ne vor indica necesitatea considerării unui anumit număr de termeni.

Să considerăm că pentru fiecare din variabilele $X_1, X_2, \ldots X_n$ am efectuat cîte o serie de determinări (observații sau măsurări). Prin urmare, pentru fiecare X_{ij} putem scrie.

$$X_{ij} = X_{ij} - \bar{x}_j + \bar{x}_j = x_{ij} + \bar{x}_j; \quad i = \overline{1, n}; \quad j = \overline{1, m} \qquad (3.89)$$

unde X_{ij} este măsurătoarea absolută, iar x_{ij} este măsurătoarea relativă în raport cu valoarea medie $\bar{x}_j$. Cu alte cuvinte x_{ij} reprezintă diferențele dintre măsurătorile absolute și media tuturor măsurătorilor de acest fel dintr-o serie determinată. Aici m este numărul de măsurători pentru o singură variabilă, iar n reprezintă numărul variabilelor în cauză.

Dacă înlocuim pe (3.89) în (3.88), obținem

$$x_{i1} + \bar{x}_1 = a_1 + b_2(x_{i2} + \bar{x}_2) + b_3(x_{i3} + \bar{x}_3) + \ldots + b_n(x_{in} + \bar{x}_n) \qquad (3.90)$$

de unde, ținînd seamă de faptul că

$$\bar{x}_1 - a_1 - b_2\bar{x}_2 - b_3\bar{x}_3 - \ldots - b_n\bar{x}_n = 0, \qquad (3.91)$$

avem mai departe

$$x_{i1} = b_2x_{i2} + b_3x_{i3} + \ldots + b_nx_{in}; \quad i = \overline{1, n} \qquad (3.92)$$

Scanned with OKEN Scanner

Ecuația (3.92) ne arată că lucrăm cu măsurătorile raportate la valoarea medie corespunzătoare, ecuațiile de acest tip au cu un termen mai puțin decît ecuațiile de forma (3.88) iar coeficienții b_k se determină cu mai mare ușurință atunci cînd folosim metoda celor mai mici pătrate. Dealtfel, dacă avem valorile medii $\bar{x}_j$ și coeficienții b_k din ecuația (3.91) se poate determina și constanta a_1.

Pentru rezolvarea sistemului (3.92) prin metoda celor mai mici pătrate, trebuie ca suma

$$S = \sum_{i=1}^{m} (b_2 x_{i2} + b_3 x_{i3} + \ldots + b_n x_{in} - x_{i1})^2 \tag{3.93}$$

să fie minimă, fapt care este condiționat de anularea derivatelor parțiale ale lui S în raport cu b_k, adică să avem

$$\frac{\partial S}{\partial b_2} = \frac{\partial S}{\partial b_3} = \ldots = \frac{\partial S}{\partial b_n} = 0 \tag{3.93a}$$

Prin urmare, ținînd seamă de (3.93a), din (3.93), avem

$$\sum_{i=1}^{m} 2x_{i2}(b_2 x_{i2} + b_3 x_{i3} + \ldots + b_n x_{in} - x_{i1}) = 0$$

$$\sum_{i=1}^{m} 2x_{i3}(b_2 x_{i2} + b_3 x_{i3} + \ldots + b_n x_{in} - x_{i1}) = 0$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\sum_{i=1}^{m} 2x_{in}(b_2 x_{i2} + b_3 x_{i3} + \ldots + b_n x_{in} - x_{i1}) = 0$$

sau, simplificînd cu 2, efectuînd înmulțirile și însumările, avem

$$\left.\begin{aligned} b_2\Sigma x_{i2}^2 + b_3\Sigma x_{i2}x_{i3} + \ldots + b_n\Sigma x_{i2}x_{in} &= \Sigma x_{i2}x_{i1} \\ b_2\Sigma x_{i2}x_{i3} + b_3\Sigma x_{i3}^2 + \ldots + b_n\Sigma x_{i3}x_{in} &= \Sigma x_{i3}x_{i1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ b_2\Sigma x_{i2}x_{in} + b_3\Sigma x_{i3}x_{in} + \ldots + b_n\Sigma x_{in}^2 &= \Sigma x_{in}x_{i1} \end{aligned}\right\} \tag{3.94}$$

Sistemul (3.94) îl supunem la următoarele transformări:

— împărțim fiecare ecuație prin n,

— sumele pătratelor le vom înlocui cu dispersiile corespunzătoare σ^2 deoarece știm că

$$\frac{\sum_{i=1}^{n} x_{ij}^2}{n} = \sigma_j^2 \tag{3.95}$$

Scanned with OKEN Scanner

— sumele dublelor produse le vom împărți prin perechi corespunzătoare de medii ale pătratelor abaterilor și vom obține coeficientul de corelație simplă care corespunde legăturii dintre cele două variabile considerate, adică dacă $j \neq k$, avem

$$\frac{\sum_{i=1}^{n} x_{ij}x_{ik}}{\sigma_j\sigma_k} \cdot \sigma_i\sigma_k = r_{jk}\sigma_j\sigma_k \tag{3.96}$$

În modul acesta, sistemul (3.94) se poate scrie

$$\left.\begin{aligned} b_2\sigma_2^2 + b_3\sigma_2\sigma_3 r_{23} + \ldots + b_n\sigma_2\sigma_n r_{2n} &= \sigma_2\sigma_1 r_{21} \\ b_2\sigma_2\sigma_3 r_{23} + b_3\sigma_3^2 + \ldots + b_n\sigma_3\sigma_n r_{3n} &= \sigma_3\sigma_1 r_{31} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ b_2\sigma_2\sigma_n r_{2n} + b_3\sigma_3\sigma_n r_{3n} + \ldots + b_n\sigma_n^2 &= \sigma_n\sigma_1 r_{n1} \end{aligned}\right\} \tag{3.97}$$

Să împărțim acum prima ecuație cu σ_2, a doua cu σ_3, ... a n-a cu σ_n; astfel că sistemul (3.97) devine

$$\left.\begin{aligned} b_2\sigma_2 + b_3\sigma_3 r_{23} + \ldots + b_n\sigma_n r_{2n} &= \sigma_1 r_{21} \\ b_2\sigma_2 r_{23} + b_3\sigma_3 + \ldots + b_n\sigma_n r_{3n} &= \sigma_1 r_{31} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ b_2\sigma_2 r_{2n} + b_3\sigma_3 r_{3n} + \ldots + b_n\sigma_n &= \sigma_1 r_{n1} \end{aligned}\right\} \tag{3.98}$$

Să ne referim acum la formula (3.75) care definește raportul de corelație

$$\eta^2 = \frac{\sigma^2(\bar{y}_x)}{\sigma_y^2}$$

și reamintim că atunci cînd regresia este liniară, coeficientul de corelație este identic cu raportul de corelație. Prin urmare, notînd cu R coeficientul de corelație obținut în condițiile specificate mai sus putem scrie

$$R^2\sigma_y^2 = \sigma^2(\bar{y}_x) = \frac{1}{m}\sum_{i=1}^{m}(\bar{y}_{xi} - \bar{y})^2 \tag{3.99}$$

unde am utilizat notațiile

$$\sigma_y^2 = \sigma_{x1}^2 = \frac{\sum_{i=1}^{m} x_{1i}^2}{n} = \sigma_1^2, \tag{3.100}$$

$$\bar{y}_{xi} - \bar{y} = x_{1i}, \tag{3.101}$$

deoarece aici am considerat măsurătorile raportate la valoarea lor medie. Astfel, x_{ij} reprezintă chiar abaterile valorilor individuale față de valoarea medie corespunzătoare.

În baza acestor considerente putem scrie mai departe

$$R^2\sigma_y^2 = \frac{1}{m}\sum_{i=1}^{m} x_{1i}^2 = \frac{1}{m}\sum_{i=1}^{m} x_{1i}\cdot x_{1i} \tag{3.102}$$

sau, ținînd seamă de (3.92), avem

$$R^2\sigma_y^2 = \frac{1}{m}\sum_{i=1}^{m} x_{1i}\,(b_2x_{i2} + b_3x_{i3} + \ldots + b_nx_{in}) =$$

$$= \frac{1}{m}\,[b_2\Sigma x_{i1}x_{i2} + b_3\Sigma x_{i1}x_{i3} + \ldots + b_n\,\Sigma x_{i1}x_{in}] \tag{3.103}$$

și, în continuare, avînd în vedere relația (3.96) și (3.100), obținem

$$R^2\sigma_y^2 = b_2\sigma_y\sigma_2r_{21} + b_3\sigma_y\sigma_3r_{31} + \ldots + b_n\sigma_y\sigma_nr_{n1}$$

de unde, după împărțirea cu σ_y, mai rămîne

$$R^2\sigma_y = b_2\sigma_2r_{21} + b_3\sigma_3r_{31} + \ldots + b_n\sigma_nr_{n1} = R^2\sigma_1 \tag{3.104}$$

Să comparăm ecuația (3.104) cu sistemul (3.98). În acest scop din (3.103) și (3.94), avem

$$R^2\sigma_y^2 = \frac{1}{m}\,(b_2^2\Sigma x_{i2}^2 + b_2b_3\Sigma x_{i2}x_{13} + \ldots + b_2b_n\Sigma x_{i2}x_{in}) +$$

$$+ \frac{1}{m}\,(b_2b_3\Sigma x_{i2}x_{i3} + b_3^2\Sigma x_{i3}^2 + \ldots + b_3b_n\Sigma x_{i3}x_{in}) + \ldots$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$+ \frac{1}{m}\,(b_2b_n\Sigma x_{i2}x_{in} + b_3b_n\Sigma x_{i3}x_{in} + \ldots + b_n^2\Sigma x_{in}^2) = b_2^2\sigma_2^2 +$$

$$+ b_2b_3\sigma_2\sigma_3r_{23} + \ldots + b_2b_n\sigma_2\sigma_nr_{2n} + \ldots + b_3^2\sigma_3^2 + \ldots$$

$$+ b_3b_n\sigma_3\sigma_nr_{3n} + \ldots = b_2\sigma_2(b_2\sigma_2 + b_3\sigma_3r_{23} + \ldots + b_n\sigma_nr_{2n}) +$$

$$+ b_3\sigma_3(b_2\sigma_2r_{23} + b_3\sigma_3 + \ldots + b_n\sigma_nr_{3n}) + \ldots$$

Se observă că, în expresia de mai sus, valorile parantezelor sînt date de (3.98) și ținînd seamă că $\sigma_1 = \sigma_y$, avem

$$R^2\sigma_y^2 = b_2\sigma_2\sigma_1r_{21} + b_3\sigma_3\sigma_1r_{31} + \ldots + b_n\sigma_n\sigma_1r_{n1}$$

unde notînd cu $\sigma_1 = \sigma_y$, obținem imediat relația (3.104). Prin urmare ecuația (3.104) este o combinație a ecuațiilor (3.98). În modul acesta, prin adăugarea ecuației (3.104) la sistemul (3.98) se obține un sistem

al cărui determinant caracteristic format din coeficienții respectivi plus coloana termenilor liberi este nul. Deci

$$\begin{vmatrix} \sigma_2 & \sigma_3 r_{23} & \dots & \sigma_n r_{2n} & \sigma_1 r_{21} \\ \sigma_2 r_{23} & \sigma_3 & \dots & \sigma_n r_{3n} & \sigma_1 r_{31} \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_2 r_{2n} & \sigma_3 r_{3n} & \dots & \sigma_n & \sigma_1 r_{n1} \\ \sigma_2 r_{21} & \sigma_3 r_{31} & \dots & \sigma_n r_{n1} & \sigma_1 R^2 \end{vmatrix} = 0$$

Dacă scoatem factori comuni : $\sigma_1, \sigma_2, \dots, \sigma_n$, de pe coloanele respective și efectuăm simplificările corespunzătoare, rezultă

$$\begin{vmatrix} 1 & r_{23} & \dots & r_{2n} & r_{21} \\ r_{23} & 1 & \dots & r_{3n} & r_{31} \\ \dots & \dots & \dots & \dots & \dots \\ r_{2n} & r_{3n} & \dots & 1 & r_{n1} \\ r_{21} & r_{31} & \dots & r_{n1} & R^2 \end{vmatrix} = 0 \qquad (3.105)$$

Determinantul (3.105) îl dezvoltăm după elementele din ultima coloană, aici vom avea semnul plus în fața lui R^2 și a minorului său. În modul acesta avem :

$$R^2 = -\frac{\begin{vmatrix} 1 & r_{23} & \dots & r_{2n} & r_{21} \\ r_{23} & 1 & \dots & r_{3n} & r_{31} \\ \dots & \dots & \dots & \dots & \dots \\ r_{2n} & r_{3n} & \dots & 1 & r_{n1} \\ r_{21} & r_{31} & \dots & r_{n1} & 0 \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} & \dots & r_{2n} \\ r_{23} & 1 & \dots & r_{3n} \\ \dots & \dots & \dots & \dots \\ r_{2n} & r_{3n} & \dots & 1 \end{vmatrix}} \qquad (3.106)$$

În cazul particular cînd $n = 2$, rezultă

$$R^2 = -\frac{\begin{vmatrix} 1 & r_{23} & r_{21} \\ r_{23} & 1 & r_{31} \\ r_{21} & r_{31} & 0 \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} \\ r_{23} & 1 \end{vmatrix}} = \frac{r_{21}^2 + r_{31}^2 - 2r_{21}r_{23}r_{31}}{1 - r_{23}^2} \qquad (3.107)$$

Scanned with OKEN Scanner

Mărimea R se numește *coeficient de corelație multiplă* și exprimă, în cazul nostru, intensitatea corelației dintre variabila X_1 și celelalte variabile: $X_2, X_3, \ldots, X_n$, de aceea, el se mai notează

$$R_{x_1, x_2, x_3 \ldots x_n}.$$

Din considerațiile prezentate mai sus, rezultă că pentru calculul coeficientului de corelație multiplă, R, trebuie ca, pentru seria de variabile obținute experimental: $x_1, x_2, \ldots, x_n$, să calculăm în mod obișnuit coeficienții de corelație $r_{x_1x_2}, r_{x_1x_3}, \ldots r_{x_{n-1}x_n}$, care sînt numiți coeficienți de corelație parțială și apoi, utilizînd formula (3.106), să-l calculăm pe R.

În cazul cînd toți coeficienții de corelație parțială ai lui X_1 și celelalte variabile au valorile apropiate de ± 1, înseamnă că între aceste mărimi există o dependență liniară. În astfel de cazuri se poate scrie sistemul (3.92), se aplică metoda celor mai mici pătrate pentru determinarea coeficienților $b_2, b_3, \ldots, b_n$. Apoi, pentru fiecare serie de valori $x_{1k}, x_{2k}, \ldots, x_{nk}$ se calculează valoarea lui y_{kc} astfel că, luînd y_{kc} în locul lui $\bar{y}_x$ definit mai înainte, putem scrie pentru raportul de corelație

$$\eta^2 = \frac{\sum_{k=1}^{n} (y_{kc} - \bar{y})^2}{\sum_{k=1}^{n} (y_k - \bar{y})^2} \tag{3.108}$$

În relația (3.108) adunăm și scădem, la numărător, cantitatea pe care o avem la numitorul respectiv, adică:

$$\eta^2 = \frac{\Sigma(y_k - \bar{y})^2 - \Sigma(y_k - \bar{y})^2 + \Sigma(y_{kc} - \bar{y})^2}{\Sigma(y_k - \bar{y})^2} = 1 - \frac{\Sigma(y_k - \bar{y})^2 - \Sigma(y_{kc} - \bar{y})^2}{\Sigma(y_k - \bar{y})^2}$$

sau

$$\eta^2 = 1 - \frac{\Sigma(y_k - y_{kc})^2 + 2\Sigma(y_{kc} - \bar{y})(y_k - y_{kc})}{\Sigma(y_k - \bar{y})^2} \tag{3.109}$$

Cel de-al doilea termen de la numărătorul lui (3.109) se poate neglija deoarece prima jumătate a diferențelor $y_{kc} - \bar{y}$ au, aproape întotdeauna, un anumit semn, iar cealaltă jumătate vor fi de semn opus. Neglijarea termenului respectiv este cu atît mai mult justificată cu cît el ar urma să fie adăugat la o sumă de pătrate care poate fi destul de mare. Din acest motiv, din (3.109) putem scrie

$$\eta^2 = 1 - \frac{\sum_{k=1}^{n} (y_k - y_{kc})^2}{\sum_{k=1}^{n} (y_k - \bar{y})^2} \tag{3.110}$$

Scanned with OKEN Scanner

Dacă ținem seamă de existența relațiilor

$$S^2 = \frac{\Sigma(y_k - y'_{kc})^2}{n}, \qquad \sigma_y^2 = \frac{\Sigma(y_k - \bar{y})^2}{n} \tag{3.111}$$

atunci

$$\eta^2 = 1 - \left(\frac{S_y}{\sigma_y}\right)^2 \tag{3.112}$$

Cantitatea η^2 astfel definită se mai numește *coeficient de determinare* iar cantitatea $(S_y/\sigma_y)^2$ este *coeficientul de înstrăinare*. Prin urmare, coeficientul de înstrăinare este definit prin raportul dintre deviația standard S_y a mărimilor y_k raportată la curba obținută prin metoda celor mai mici pătrate (y_{kc}) și deviația standard a acelorași mărimi y_k raportată la valoarea lor medie $\bar{y}$.

Din relația (3.112) se vede că cu cît este mai mare gradul de corelație, adică cu cît dependența dintre y și x, de exemplu, este mai mare, cu atît S_y/σ_y va fi mai mic.

Ecuația (3.112) se mai poate scrie și sub forma:

$$\eta^2 + \left(\frac{S_y}{\sigma_y}\right)^2 = 1 \tag{3.113}$$

fapt care ne arată că locul geometric al funcției

$$f(\eta) = S_y/\sigma_y \tag{3.114}$$

este un cerc.

Dacă utilizăm raportul de corelație sub forma definită de relațiile (3.108) sau (3.110), unde intervin mărimile y_{kc}, adică unde utilizăm valorile calculate ale lui y, atunci aceleași relații pot fi utilizate și pentru corelația neliniară sau multiplă. De altfel, relația (3.107) este dedusă în ipoteza existenței unei regresii liniare, fapt care nu poate fi totdeauna adevărat. De aceea relațiile (3.108) sau (3.110) pot fi de mare utilitate practică, bineînțeles reținînd ideea că la deducerea lui (3.110) s-a neglijat un termen care de fapt este destul de mic.

Pentru corelația parțială și multiplă despre care am vorbit mai sus, am spus că vom determina coeficienții r_{x_1, x_2}, $r_{x_1, x}$ etc. și apoi, dacă toți aceștia indică existența unei regresii liniare, putem aplica formula (3.106). Dar dacă nu sîntem siguri de existența unei corelații liniare, atunci putem folosi relația (3.110) scrisă sub forma

$$\eta = \sqrt{1 - \frac{\Sigma(y_k - y'_{kc})^2}{\Sigma(y_k - \bar{y})^2}} \tag{3.115}$$

unde y'_{kc} este valoarea calculată pentru seria de mărimi x_{ik} după ce, de fapt, prin metoda celor mai mici pătrate s-a obținut o formulă

Scanned with OKEN Scanner

empirică. Evident, aici intervine dificultatea mare de a şti să alegem forma cea mai potrivită pentru dependenţa probabil existentă în cazul considerat. Această problemă va fi discutată în capitolul următor.

În cazul cînd utilizăm formule empirice liniare raportul de corelaţie este identic cu coeficientul de corelaţie. Prin urmare, din (3.108), pentru coeficientul de corelaţie multiplă putem scrie

$$R_{y,\,x_2,\,x_3,\,\ldots,\,x_n} = \sqrt{1 - \frac{\Sigma(y_k - \bar{y}_{x_2,\,x_3,\,\ldots,\,x_n})^2}{\Sigma(y_k - \bar{y})^2}} \tag{3.116}$$

unde citim: coeficientul de corelaţie dintre y şi variabilele $x_2, x_3, \ldots x_n$ etc. se exprimă în funcţie de valorile calculate $\bar{y}_{x_2,\,x_3,\,x_4,\,\ldots,\,x_n}$ pentru calculul cărora s-a folosit o anumită formulă empirică determinată.

*
* *

Am specificat mai înainte semnificaţia şi importanţa coeficientului de corelaţie pentru cazul cînd între X_k şi Y_k există o regresie liniară. Într-adevăr, dacă ne referim la expresia matematică a acestui coeficient [vezi (3.41)]

$$r_{y/x} = \frac{\Sigma(x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\Sigma(\bar{x}_k - \bar{x})^2 \Sigma(y_k - \bar{y})^2}} \tag{3.117}$$

în cazul regresiei liniare, se pot efectua următoarele înlocuiri

$$y_k = ax_k + b\,; \text{ sau } \bar{y} = \frac{1}{n}\Sigma y_k = \frac{a}{n}\Sigma x_k + b\,; \; \bar{y} = a\bar{x} + b$$

În aceste condiţii, formula (3.117) se scrie

$$r_{y/x} = \frac{\Sigma(x_k - \bar{x})(y_k - a\bar{x} - b)}{\sqrt{\Sigma(x_k - \bar{x})^2 \Sigma(y_k - a\bar{x} - b)^2}} \tag{3.118}$$

Prin urmare, dacă y_k se va găsi pe linia de regresie, vom avea $y_k = ax_k + b$ pe care o înlocuim în (3.118) şi obţinem

$$r_{y/x} = \frac{a\Sigma(x_k - \bar{x})(x_k - \bar{x})}{+a\sqrt{\Sigma(x_k - \bar{x})^2 \Sigma(x_k - \bar{x})^2}} = \pm 1 \tag{3.119}$$

adică pentru o corelaţie perfect liniară r va avea valoarea ± 1, rezultat care a fost găsit şi mai înainte (vezi formula 3.51).

Dacă modul de a examina problemele de corelaţie, aşa cum a fost prezentat mai sus, ar fi extins şi la regresiile curbilinii, am putea determina şi pentru ele un „raport de corelaţie" asemănător. În acest sens, pentru o „parabolă" de un ordin oarecare, avem

$$y_k = a_0 + a_1 x_k + a_2 x_k^2 + a_3 x_k^3 + \ldots + a_n x_k^n \tag{3.120}$$

Scanned with OKEN Scanner

unde, pe cale experimentală, dispunem de seriile de valori x_k, y_k. Apoi, prin metoda celor mai mici pătrate, determinăm constantele a_0, a_1, a_2, ..., a_n iar formula (3.114) devine

$$\eta^2_{y/x} = 1 - \frac{\sum_{k=1}^{n} (y_k - a_0 - a_1 x_k - a_2 x_k^2 - \ldots - a_n x_k^n)^2}{\sum_{k=1}^{n} (y_k - \bar{y})^2} \qquad (3.121)$$

care definește „raportul de corelație curbilinie".

Aplicație practică

Pentru a exemplifica modul de aplicare practică a considerațiilor expuse mai sus, am utilizat tabela nr. 89 din lucrarea lui Romanovski (Op. cit. pg. 618) în care sînt date cotele apelor unui rîu măsurate la același moment în trei stații diferite. Aceste date le reproducem în primele trei coloane ale tabelei 3.6 unde cele trei cote măsurate sînt reprezentate prin y, x_1 și respectiv x_2. Celelalte coloane au semnificațiile specificate și reprezintă calculele intermediare necesare pentru determinarea coeficienților de corelație parțială: r_{y/x_1}, r_{y/x_2} și r_{x_1/x_2} precum și pentru calcularea coeficientului de corelație multiplă R.

a) *Utilizarea formulelor* (3.43) și (3.114)

Ultima linie a tabelei 3.6 reprezintă sumele coloanelor respective, iar valorile medii sînt:

$$\bar{y} = \frac{1235}{61} = 20{,}25 \simeq 20, \qquad \bar{x}_1 = \frac{1340}{61} = 22, \qquad \bar{x}_2 = \frac{3025}{61} \simeq 50$$

Rotunjirile valorilor medii au fost efectuate cu scopul de-a avea cît mai puține cifre și a reduce la minim volumul de calcule și spațiul corespunzător.

Aplicînd formula (3.43) pentru determinarea celor trei coeficienți de corelație parțială, avem

$$r_{y/x_1} = \frac{8870}{\sqrt{13975 \times 6514}} = 0{,}9297, \qquad r_{y/x_2} = \frac{6950}{\sqrt{13975 \times 3975}} = 0{,}9325$$

$$r_{x_1/x_2} = \frac{4825}{\sqrt{6514 \times 3975}} = 0{,}9481$$

Pentru calcularea coeficientului de corelație multiplă, R, am utilizat formula (3.107) care, în cazul notațiilor noastre, are forma

$$R^2_{y,\,x_1,\,x} = \frac{(r_{y/x_1})^2 + (r_{y/x_2})^2 - 2 r_{y/x_1} r_{y/x_2} r_{x_1/x_2}}{1 - (r_{x_1/x_2})^2}$$

sau, efectuînd înlocuirile corespunzătoare, avem mai departe

$$R^2 = \frac{(0,9297)^2 + (0,9325)^2 - 2(0,9297)(0,9325(0,9481)}{1 - (0,9481)^2} = \frac{0,08997}{0,10111}$$

adică $R = 0,9433$.

b) *Utilizarea formulelor (3.115) și (3.121)*

În cazul cînd avem deja calculate diferențele: $y_k - \bar{y}$, $x_1 - \bar{x}_1$ și $x_2 - \bar{x}_2$ (vezi coloanele 5, 6 și 7 din tabela 3,6), putem scrie relația (3.92) astfel că vom avea sistemul de ecuații de condiții de forma

$$a_2(x_2 - \bar{x}_2)_i + a_1(x_1 - \bar{x}_1)_i = (y - \bar{y})_i; \qquad i = \overline{1,22}. \tag{3.122}$$

Pentru rezolvarea sistemului (3.122) și determinarea necunoscutelor a_2 și a_1 vom utiliza metoda celor mai mici pătrate (vezi I. Todoran, 1976, pg. 133). În modul acesta rezultă următorul sistem de ecuații normale

$$a_2\Sigma(x_2 - \bar{x}_2)_i^2 + a_1\Sigma(x_2 - \bar{x}_2)(x_1 - \bar{x}_1) = \Sigma(y - \bar{y})(x_2 - \bar{x}_2)$$

$$a_2\Sigma(x_2 - \bar{x}_2)(x_1 - \bar{x}_1) + a_1\Sigma(x_1 - \bar{x}_1)^2 = \Sigma(y - \bar{y})(x_1 - \bar{x}_1)$$

unde înlocuim valorile numerice corespunzătoare din tabela 3.6 și avem

$$3975\ a_2 + 4825\ a_1 = 6950$$

$$4825\ a_2 + 6514\ a_1 = 8870$$

de unde rezultă

$$a_2 = 0,9476; \qquad a_1 = 0,6606$$

Cu ajutorul acestor constante astfel determinate revenim la coloanele 5, 6 și 7 din tabela 3.6 și întocmim tabela 3.7 în care prima coloană conține mărimile determinate experimental $Y_k = y_k - \bar{y}$ iar în coloanele 2 și 3 apar mărimile necesare pentu a calcula valorile Y_{ck} definite de

$$Y_{ck} = a_2(x_2 - \bar{x}_2)_k + a_1(x_1 - \bar{x}_1)_k$$

care sînt reproduse în coloana 4 a tabelei 3.7. În modul acesta putem face înlocuirile în formula (3.121) pe care am scris-o sub forma

$$\eta_{y/x} = \sqrt{1 - \frac{\sum_{k=1}^{n} n_k[y_k - a_2(x_2 - \bar{x}_2)_k - a_1(x_1 - \bar{x}_1)_k]}{\sum_{k=1}^{n} n_k(y_k - \bar{y})^2}} \tag{3.123}$$

Numărătorul fracției din formula (3.123) este definit de mărimile din penultima coloană a tabelei 3.7 și are valoarea 1482 iar numitorul

Scanned with OKEN Scanner

respectiv are valoarea $\sum_{k=1}^{n}(y_k - \bar{y})^2 = 13975$ (vezi coloana 8 din tabela 3.6).

Prin urmare,

$$\eta_{y/x} = \sqrt{1 - \frac{1482}{13975}} = 0{,}9454.$$

Comparația dintre această ultimă valoare $\eta_{y/x} = 0{,}9454$ și mărimea $R = 0{,}9433$ obținută mai sus (vezi subpunctul precedent) ne duce la concluzia că cele două procedee au dat același rezultat, așa cum era de așteptat în cazul corelațiilor liniare.

Tabela 3.6

	x_1	x_2	n	$y-\bar{y}$	$x_1-\bar{x}_1$	$x_2-\bar{x}_2$	$(y-\bar{y})^2$	$(x_1-\bar{x}_1)^2$	$(x_2-\bar{x}_2)^2$	$(y-\bar{y})(x_1-\bar{x}_1)$	$(y-\bar{y})(x_2-\bar{x}_2)$	$(x_1-\bar{x}_1)(x_2-\bar{x}_2)$
5	10	40	8	−15	−12	−10	225	144	100	180	150	120
5	10	45	1	−15	−12	− 5	225	144	25	180	75	60
5	15	45	7	−15	− 7	− 5	225	49	25	105	75	35
10	10	40	1	−10	−12	−10	100	144	100	120	100	120
10	15	40	4	−10	− 7	−10	100	49	100	70	100	70
10	15	45	7	−10	− 7	− 5	100	49	25	70	50	35
10	20	45	2	−10	−22	− 5	100	4	25	20	50	10
15	20	45	1	− 5	− 2	− 5	25	4	25	10	25	10
15	20	50	2	− 5	− 2	0	25	4	0	10	0	0
20	20	50	2	0	− 2	0	00	4	0	0	0	0
20	25	50	4	0	3	0	0	0	0	0	0	0
25	25	55	3	5	3	5	25	9	25	15	25	15
30	25	55	3	10	3	5	100	9	25	30	50	15
35	30	55	2	15	8	5	225	64	25	120	75	40
40	30	55	3	20	8	5	400	64	25	160	100	40
45	30	60	2	25	8	10	625	64	100	200	250	80
45	35	60	4	25	13	10	625	169	100	325	250	130
45	40	65	1	25	18	15	625	324	225	450	375	270
45	45	65	1	25	23	15	625	529	225	575	375	345
45	45	70	1	25	23	20	625	529	400	575	500	460
45	50	60	1	25	28	10	625	784	100	700	250	280
45	50	70	1	25	28	20	625	784	400	700	500	560
1235	1340	3025	61				13975	6514	3975	8870	6950	4825

Scanned with OKEN Scanner

Tabela 3.7

$Y_k = y_k - \bar{y}$	$a_1(x_1 - \bar{x}_1)$	$a_2(x_2 - \bar{x}_2)$	$(Y_c)_k$	$(Y - Y_c)_k$	$(Y - Y_c)_k^2$	n
−15	−8	−10	−18	3	9	8
−15	−8	− 5	−13	− 2	4	1
−15	−5	− 5	−10	− 5	25	7
−10	−8	−10	−18	+ 8	64	1
−10	−5	−10	−15	+ 5	25	4
−10	−5	− 5	−10	0	0	7
−10	−1	− 5	− 6	− 4	16	2
− 5	−1	− 5	− 6	+ 1	1	1
− 5	−1	0	− 1	− 4	16	2
0	−1	0	− 1	− 1	1	2
0	+2	0	+ 2	− 2	4	4
5	2	5	7	− 2	4	3
10	2	5	7	3	9	3
15	5	5	10	5	25	2
20	5	5	10	10	100	3
25	5	10	15	10	100	2
25	9	10	19	6	36	4
25	12	14	26	− 1	1	1
25	15	14	29	− 5	16	1
25	15	19	34	− 9	81	1
25	18	10	28	− 3	9	1
25	18	19	37	−12	144	1

3.12. Funcția de corelație

Dacă ne referim la formulele (3.41), (3.41a) și (3.72), se constată că suma $\Sigma x_k y_k$ este termenul care cauzează variația în mărimea coeficientului de corelație. Această cantitate am mai întîlnit-o și în formula (3.18) unde avem

$$\nu_{11} = \frac{1}{n} \sum_{k=1}^{n} x_k y_k \tag{3.15}$$

sau, în cazul cînd $\bar{x} \neq 0$, $\bar{y} \neq 0$, formula (3.18) era

$$\mu_{11} = \nu_{11} - \bar{x}\bar{y} = \frac{1}{n} \sum_{k=1}^{n} x_k y_k - \bar{x}\bar{y} \tag{3.18}$$

care mai poate fi scrisă

$$\mu_{11} = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x}_k)(y_k - \bar{y}) \tag{3.124}$$

sau, utilizînd notațiile din statistica matematică,

$$\mu_{11} = M[x - M(x)][y - M(y)] \tag{3.124a}$$

Scanned with OKEN Scanner

unde μ_{11} se numeşte momentul centrat de ordinul doi al mărimilor accidentale x_k şi y_k, iar $M(x)=\frac{1}{n}\sum_{k=1}^{n} x_k$ este aşteptarea sau speranţa matematică sau valoarea medie a mărimilor x_k.

Avînd în vedere particularităţile coeficientului de corelaţie, $r = 0$ atunci cînd între x_k şi y_k nu există nici o corelaţie, adică $\mu_{11} = 0$ cînd x_k şi y_k sînt independente, momentul centrat de ordinul doi al mărimilor respective se numeşte *moment de corelaţie.*

În cazul seriilor dinamice x_k şi y_k sînt funcţii de timp, adică trebuie să scriem $x(t_k)$ şi $y(t_k)$ atunci cînd cele două variabile sînt socotite la acelaşi moment t_k — serii sincrone — sau $x(t_k)$ şi $y(t_k + \Delta t)$ pentru seriile asincrone.

În cazul seriilor sincrone, momentul de corelaţie va avea expresia

$$\mu_{11}(t_k,\ t_v) = M\{\{x(t_k) - M[x(t_k)]\}\cdot\{y(t_v) - M[y(t_v)]\}\} \qquad (3.125)$$

care, pentru corelaţia internă — autocorelaţie — devine

$$\mu_{11}(t_k,\ t_v) = M\{\{x(t_k) - M[x(t_k)]\}\cdot\{x(t_v) - M[x(t_v)]\}\} \qquad (3.126)$$

Dacă se fixează unul din momente, de exemplu t_k, iar t_v se consideră variabil, atunci variaţia respectivă va cauza o variabilitate a mărimii μ_{11}.

În cazul cînd considerăm o diferenţă constantă $t_k - t_v = \tau$, cantitatea $\mu_{11}(t_k,\ t_v)$ poate deveni o funcţie atît de timpul t cît şi de τ. Această funcţie se numeşte *funcţie de corelaţie* $R_{xy}(t, \tau)$ sau, în cazul unei corelaţii interne; *funcţie de autocorelaţie* $R_{xx}(t, \tau)$. Prin urmare,

$$R_{xy}(t, \tau) = R_{xy}(t_k t_v) = \mu_{11} = M\{\{x(t_k) - M[x(t)]\}\cdot\{y(t-\tau) - M[y(t-\sigma)]\}\} \qquad (3.127)$$

În cazul unei aceleiaşi serii de date avem funcţia de autocorelaţie definită prin relaţia.

$$R_{xx}(t, \tau) = M\{\{x(t) - M[x(t)]\}\cdot\{x(t-\tau) - M[x(t-\tau)]\}\} \qquad (3.128)$$

Dacă $\tau = 0$, funcţia de autocorelaţie este chiar dispersia, adică

$$R_{xx}(t_k,\ t_k) = R_{xx}(t,\ 0) = M\{\{x(t) - M[x(t)]\}^2\} = D_x^2(t), \qquad (3.129)$$

ceea ce înseamnă că, în cadrul seriilor dinamice, dispersia este o funcţie de timp.

Dacă ne referim acum la nişte serii în care valorile medii luate pe anumite intervale de timp sînt egale, sau aproape egale cu media generală a ansamblului de mărimi considerate, se spune că avem de-a face cu *procese stochastice ergodice.* În aceste condiţii, de la media pe anumite intervale de timp se poate trece la media pe ansamblu,

Scanned with OKEN Scanner

astfel că pentru funcția de corelație se poate considera valoarea medie definită de integrala

$$R_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)y(t+\tau)\mathrm{d}t \tag{3.130}$$

unde $2T$ este intervalul de timp pe care se efectuează medierea produsului $x(t)y(t+\tau)$, iar τ reprezintă mărimea deplasării dintre momentele la care sînt considerate cele două variabile.

Formula (3.130) este scrisă ca și cum pentru funcția de corelație din formula (3.41) am fi luat în considerare numai partea variabilă.

Dacă ne referim la o singură variabilă, atunci funcția de autocorelație va fi definită prin expresia

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t+\tau)\mathrm{d}t \tag{3.131}$$

care este aceeași cu

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t-\tau)\cdot x(t)\mathrm{d}t \tag{3.132}$$

care, pentru $\tau = 0$, devin

$$R_{xx}(0) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x^2(t)\mathrm{d}t \tag{3.133}$$

De aici se vede că funcția de autocorelație $R_{xx}(\tau)$ are valoarea maximă pentru $\tau = 0$, deoarece, în acest caz, $R_{xx}(\tau)$ tinde spre $R_{xx}(0)$ care reprezintă dispersia mărimilor $x(t)$.

În cazul cînd $|\tau| \to \infty$ mărimea funcției de corelație tinde spre zero deoarece, în acest caz, sub integrala din (3.132) vom avea cel mult produsul finit a doi termeni, care, pentru $T \to \infty$ duce la $R_{xx}(|\tau| \to \infty) = 0$.

Schematic, graficul funcției de corelație $R_{xy}(\tau)$ și al celei de autocorelație $R_{xx}(\tau)$ sînt prezentate pe figura 3.7.

După cum rezultă din formulele (3.131) și (3.132) se vede că funcția de autocorelație este o funcție pară deoarece avem

$$R_{xx}(\tau) = R_{xx}(-\tau) \tag{3.134}$$

Prin urmare, datorită structurii sale, pentru construcția funcției $R_{xx}(\tau)$ ne putem limita la valorile lui τ cuprinse între 0 și ∞. O astfel de concluzie nu este valabilă pentru funcția de corelație deoarece

$$R_{xy}(\tau) \neq R_{yx}(-\tau) \tag{3.135}$$

Scanned with OKEN Scanner

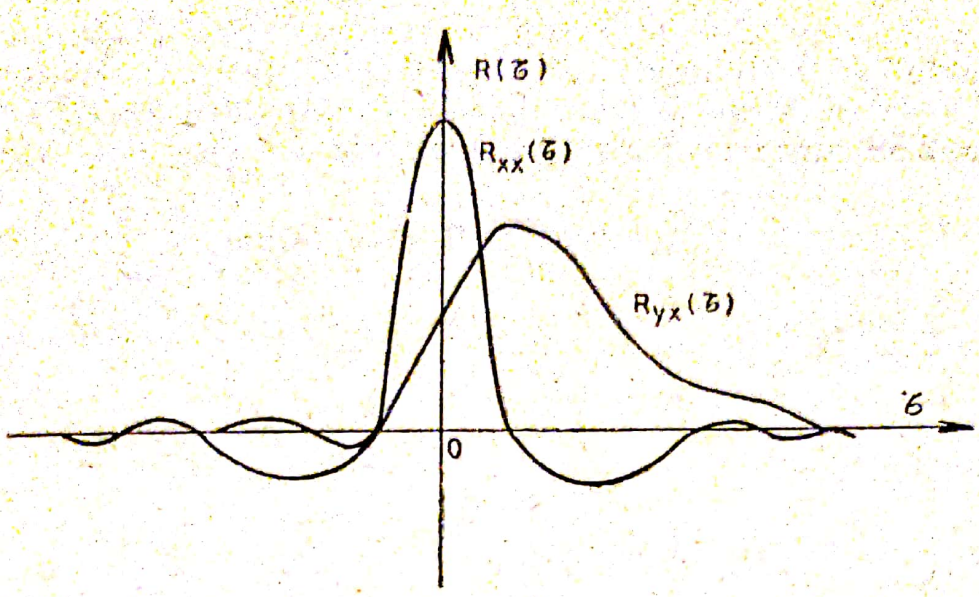


Fig. 3.7. Reprezentarea schematică a funcției de autocorelație $R_{xx}(\tau)$ și a celei de corelație $R_{xy}(\tau)$.

Cu foarte rare excepții, funcțiile de corelație pot fi considerate ca funcții continue care pot lua atît valori pozitive cît și negative din domeniile lor de definiție.

Dacă la definirea funcției de corelație am fi luat în considerare expresia (3.126), atunci în loc de (3.130) am fi avut

$$R_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \{x(t) - M[x(t)]\} \cdot \{y(t+\tau) - M[y(t)]\} dt =$$

$$= \lim \frac{1}{2T} \int_{-T}^{T} x(t)y(t+\tau)dt - M[x(t)]\ M[y(t)] \qquad (3.136)$$

deoarece se ia în considerare numărătorul expresiei (3.43a) și se are în vedere faptul că valoarea medie a unei constante este egală cu constanta respectivă.

Uneori este comod să normalizăm funcția de corelație, astfel că funcția de corelație normalizată, $R^N_{xx}(\tau)$, va fi definită prin formula

$$R^N_{xx}(\tau) = \frac{R_{xx}(\tau)}{R_{xx}(0)} \qquad (3.137)$$

ceea ce ne conduce la concluzia că mărimea funcției de autocorelație normalizată la $\tau = 0$ este egală cu unitatea, adică

$$R^N_{xx}(0) = 1 \qquad (3.138)$$

Dacă revenim asupra semnalului aleator $x(t)$ și ne reamintim considerațiile din analiza spectrală (vezi capitolul II), se trage con-

cluzia că acest semnal va fi caracterizat printr-un spectru care şi el la rîndul lui, va fi o funcţie aleatoare. De aceea, pentru scopuri practice ne putem folosi de spectrul funcţiei de autocorelaţie $R_{xx}(\tau)$ care (vezi formula 2.59) duce la densitatea spectrală

$$S_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau)\, e^{-i\omega\tau}\, d\tau \tag{3.139}$$

Dacă se face uz de formula lui Euler scrisă sub forma .

$$e^{-i\omega\tau} = \cos\omega\tau - i \sin \omega\tau$$

avem mai departe

$$S_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau) \cos \omega\tau \cdot d\tau - i \int_{-\infty}^{\infty} R_{xx}(\tau) \sin \omega\tau \cdot d\tau =$$

$$= \int_{-\infty}^{0} R_{xx}(\tau) \cos \omega\tau \cdot d\tau + \int_{0}^{\infty} R_{xx}(\tau) \cos \omega\tau \cdot d\tau -$$

$$- i \int_{-\infty}^{0} R_{xx}(\tau) \sin \omega\tau \cdot d\tau - i \int_{0}^{\infty} R_{xx}(\tau) \sin \omega\tau \cdot d .$$

Dacă ţinem seamă de faptul că

$$\int_{-\infty}^{0} R_{xx}(\tau) \cos \omega\tau \cdot d\tau = \int_{0}^{\infty} R_{xx}(\tau) \cos \omega\tau \cdot d\tau$$

şi

$$\int_{-\infty}^{0} R_{xx}(\tau) \sin \omega\tau \cdot d\tau = - \int_{0}^{\infty} R_{xx}(\tau) \sin \omega\tau \cdot d\tau$$

putem scrie

$$S_{xx}(\omega) = 2 \int_{0}^{\infty} R_{xx}(\tau) \cos \omega\tau \cdot d\tau \tag{3.140}$$

Utilizînd transformata Fourier definită prin (2.44) şi (2.45), din formula (3.139) avem

$$R_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{xx}(\omega) e^{i\omega\tau}\, d\omega \tag{3.141}$$

Formulele (3.139) și (3.141) definesc relațiile de trecere de la funcția de autocorelație la densitatea spectrală și invers, folosind în acest scop calculul integralelor care definesc transformata Fourier.

În domeniul timpului avem funcția de autocorelație care exprimă legătura dintre semnalele distanțate prin intervalul τ, în domeniul frecvențelor această legătură fiind exprimată prin funcția de densitate spectrală. Prin urmare, formulele (3.139) și (3.141) ne permit să facem trecerea de la domeniul timpului la domeniul frecvențelor și invers.

În cazul cînd este vorba de corelația dintre două serii dinamice diferite, pentru evaluarea densității spectrale $S_{xy}(\omega)$ se utilizează funcția de corelație $R_{xy}(\tau)$, astfel că

$$S_{xy}(\omega) = \int_{-\infty}^{\infty} R_{xy}(\tau) e^{-i\omega\tau} \, d\tau \tag{3.142}$$

Utilizarea formulelor (3.139) sau (3.142) pentru determinarea densității spectrale este destul de incomodă. De aceea, la evaluarea mărimilor acestei funcții pentru anumite frecvențe ω_k, ne putem folosi de însăși definiția acesteia astfel că, din (2.59), putem scrie

$$S_{xx}(\omega_k) = \frac{c_{xk}}{P} = \frac{\sqrt{a_{xk}^2 + b_{xk}^2}}{P} \tag{3.143}$$

unde a_{xk} și b_{xk} sînt coeficienții Fourier ai semnalului $x(t)$ și corespund intervalului de la 0 la P.

Pentru evaluarea funcției de corelație putem face uz de relațiile prin care a fost introdusă această funcție (vezi formulele 3.124 și 3.129).

Aplicație practică

Pentru a ilustra modul în care se poate utiliza funcția de corelație în probleme de prognoză, vom porni de la relația (3.137) unde este definită funcția de autocorelație normată și care, de altfel, este analogă cu expresia coeficientului de corelație (vezi formula 3.41).
Prin urmare

$$R_{xx}^{N}(\tau) = \frac{\frac{1}{n-\tau} \Sigma [x(t_k) - \bar{x}(t_k)] [x(t_k - \tau) - \bar{x}(t_k - \tau)]}{\frac{1}{n} \Sigma [x(t_k) - \bar{x}(t_k)]^2} \tag{3.144}$$

$(\tau = 1, 2, \ldots, n-1)$; $k = 1, 2, \ldots, n - \tau$.

În vederea utilizării practice a formulei (3.144) vom face uz de tabela 6.8 din lucrarea lui Mihoc, Bergthaller și Urseanu (1978), unde este reprodusă dinamica producției de grîu și secară a R.S.R. în

Scanned with OKEN Scanner

perioada 1951—1975. Astfel, pentru calculele pe care le vom efectua, întocmim tabela noastră 3.8a, unde în primele două coloane (reproduse după lucrarea amintită mai sus) se află timpul t exprimat în ani și respectiv producția de grîu și secară $x_0(t)$ exprimată în chintale la hectar.

Din examinarea primelor două coloane ale tabelei 3.8a se vede deja că există o tendință generală a producției de grîu și secară la hectar, există un trend cvasiliniar care poate fi utilizat pentru probleme de prognoză. Dar, în plus, există și anumite variații care par să aibă un caracter oscilator. Prin urmare, din examinarea acestor prime două coloane ale tabelei 3.8a putem desprinde următoarele două probleme importante pentru prognoză:

(a) care va fi producția probabilă de grîu și secară la hectar în anii 1975—1980, și

(b) sînt accidentale oscilațiile observate în producția respectivă, sau există vreo tendință de variație periodică?

Asupra acestor două probleme vom insista în cele ce urmează, folosind în acest scop funcția de autocorelație normată.

După cum se constată, numitorul de la fracția formulei (3.144) este o mărime constantă în care intră toate mărimile din coloana a 2-a a tabelei 3.8a, mărimea medie a acestora fiind definită prin relația

$$\bar{x}_0(t) = \frac{\sum_1^{25} x_0(t_k)}{25} = \frac{12{,}5 + 10{,}7 + \ldots + 20{,}8 + 20{,}6}{25} = 15{,}2$$

Efectuînd diferența dintre valorile conținute în coloana a 2-a și mărimea medie 15,2, obținem diferențele $x_0(t_k) - \bar{x}_0(t_k)$ care sînt trecute în coloana a 3-a, capul coloanelor fiind marcat prin $\tau + 1 = 1, 2, 3 \ldots, 25$ și reprezintă numărul de ani scurși după anul 1950.

Prin urmare, pentru calcularea numitorului fracției din formula (3.144) trebuie să calculăm suma pătratelor diferențelor din coloana a 3-a, această sumă fiind

$$\frac{1}{25}\sum_1^{25}[x_0(t_k) - \bar{x}_0(t_k)]^2 = \frac{503{,}54}{25} = 20{,}14.$$

Pe baza acestor considerații, funcția de autocorelație normată, pentru $t = 1951$, adică $\tau = 0$, are valoarea

$$R_{xx}^N(0) = \frac{20{,}14}{20{,}14} = 1{,}00$$

Pentru $\tau = 1$, adică în vederea calculării funcției de corelație pentru cazul cînd $t = 1950 + 2 = 1952$, avem mărimea medie

$$\bar{x}_0(t_k) = \frac{1}{24}\sum_2^{25} x_0(t_k) = \frac{367{,}3}{24} = 15{,}3$$

care reprezintă media tuturor valorilor din coloana a 2-a unde însă am lăsat la o parte prima valoare (12,5). Procedînd în mod analog, calculînd mediile valorilor din coloana a 2-a în cazurile cînd se atribuie lui k valorile de la 3 la 25, de la 4 la 25 etc, se obţin mediile pe care le-am trecut în penultima linie a tabelei 3.8a. Efectuînd diferenţele dintre mărimile corespunzătoare ale coloanei a 2-a şi mediile din penultima linie, rezultă valorile din celelalte coloane ale tabelei 3.8a. Apoi, efectuînd produsul dintre diferenţele din coloana a 3-a şi diferenţele corespunzătoare din coloanele următoare şi însumînd, se obţin succesiv valorile numărătorului fracţiei (3.144) pe care le împărţim la valoarea constantă a numitorului (= 20,14). În modul acesta rezultă valorile funcţiei de autocorelaţie care sînt date în ultima linie a tabelei 3.8a. Aceste valori corespund anilor 1950 + numărul din capul coloanelor respective.

Pe figura 3.3 este reprezentat graficul funcţiei $x_0(t) = f(t)$, adică graficul producţiei de grîu şi secară în funcţie de timpul respectiv. Din examinarea acestui grafic se pune în evidenţă prezenţa trendului cvasi-liniar, adică se constată că, în general, producţia respectivă creşte, dar această creştere este însoţită de o serie de fluctuaţii care par să aibă un caracter accidental. Liniaritatea acestei creşteri a fost examinată de Mihoc, Bergthaller şi Urseanu (1978) dînd următoarea relaţie.

$$x_0(t) = 15{,}12 + 0{,}5218(t - 1963)$$

unde timpul t este exprimat în ani.

Făcînd abstracţie de formula de mai sus, noi am aplicat formula (3.144) direct la datele „originale" (vezi primele două coloane ale tabelei 3.8a). Astfel am calculat valorile funcţiei de autocorelaţie normată care sînt date în ultima linie a tabelei 3.8a, iar reprezentarea grafică a acestora se găseşte pe figura 3.8. De aici se pune în evidenţă faptul că funcţia de autocorelaţie normată are o descreştere de la 1 la zero, aşa cum s-a stabilit şi pe cale teoretică. Prin urmare, făcînd abstracţie de micile oscilaţii accidentale ale funcţiei respective — urmărind curba trasată cu mîna liberă printre punctele graficului — putem trage concluzia că datele din primele coloane ale tabelei 3.8a, prezintă o autocorelaţie de care trebuia să se ţină seamă atunci cînd s-a discutat prognoza producţiei de grîu şi secară pentru anii 1975—1980.

Luînd în considerare formula de mai sus, pentru anul 1980 avem

$$x_0(1980) = 15{,}12 + 0{,}5218(1980 - 1963) \simeq 24$$

care este o mărime probabilă şi care putea să oscileze între 24 + 5 şi 24—5 chintale de grîu şi secară la hectar. Aceste limite sînt provocate de anumiţi factori necunoscuţi dinainte şi de care nu putem

Anul	$x_0(t)$	1	2	3	4	5	6	7	8	9	10	11	12
1951	12,5	−2,7	−4,6	−1,1	−6,9	−5,7	−7,8	−4,1	−7,0	−3,8	−5,4	−4,4	−4,8
1952	10,7	−4,5	−0,9	−6,8	−5,4	−7,5	−3,7	−6,8	−3,4	−5,1	−4,1	−4,5	−4,9
1953	14,4	−0,8	−6,6	−5,3	−7,2	−3,4	−6,4	−3,2	−4,7	−3,8	−4,2	−4,6	−5,2
1954	8,7	−6,5	−5,1	−7,1	−3,1	−6,1	−2,8	−4,5	−3,4	−3,9	−4,3	−4,9	+1,6
1955	10,2	−5,0	−6,9	−3,0	−5,8	−2,5	−4,1	−3,2	−3,5	−4,0	−4,6	+1,9	−1,4
1956	8,4	−6,8	−2,8	−5,7	−2,2	−3,8	−2,8	−3,3	−3,6	−4,3	+2,2	−1,1	+1,9
1957	12,5	−2,7	−5,5	−2,1	−3,5	−2,5	−2,9	−3,4	−3,9	+2,5	−0,8	+2,2	−0,9
1958	9,8	−5,4	−1,9	−3,4	−2,2	−2,6	−3,0	−3,7	+2,9	−0,5	+2,5	−0,6	−2,3
1959	13,4	−1,8	−3,2	−2,1	−2,3	−2,7	−3,3	+3,1	−0,1	+2,8	−0,3	−2,0	−3,7
1960	12,1	−3,1	−1,9	−2,2	−2,4	−3,0	+3,5	+0,1	+3,2	0,0	−1,7	−3,4	+4,1
1961	13,4	−1,8	−2,0	−2,3	−2,7	+3,8	+0,5	+3,4	+0,4	−1,4	−3,1	+4,4	+5,7
1962	13,3	−1,9	−2,1	−2,6	+4,1	+0,8	+3,8	+0,6	−1,0	−2,8	+4,7	+6,0	+5,0
1963	13,2	−2,0	−2,4	+4,2	+1,1	+4,1	+1,0	−0,8	−4,4	+5,0	+6,3	+5,3	+2,7
1964	12,9	−2,3	+4,4	+1,2	+4,4	+1,3	−0,4	−2,2	+5,4	+6,6	+5,6	+3,0	+2,5
1965	19,7	+4,5	+1,4	+4,5	+1,6	−0,1	−1,8	+5,6	+7,0	+5,9	+3,3	+2,8	
1966	16,7	+1,5	+4,7	+1,7	+0,2	−1,5	+6,0	+7,2	+6,3	+3,6	+3,1		
1967	20,0	+4,8	+1,9	+0,3	−1,2	+6,3	+7,6	+6,5	+4,0	+3,4			
1968	17,2	+2,0	+0,5	−1,1	+6,6	+7,9	+6,9	+4,2	+3,8				
1969	15,8	+0,6	−0,9	+6,7	+8,2	+7,2	+4,6	+4,0					
1970	14,4	−0,8	+6,9	+8,3	+7,5	+4,9	+4,4						
1971	22,2	+7,0	+8,5	+7,6	+5,2	+4,7							
1972	23,8	+8,6	+7,8	+5,3	+5,0								
1973	23,1	+7,9	+5,5	+5,1									
1974	20,8	+5,6	+5,3										
1975	20,6	+5,4											
$\bar{x}_0(t)$		15,2	15,3	15,5	15,6	15,9	16,2	16,6	16,8	17,2	17,5	17,8	18,1
$R_{xx}(t)$		1,00	+0,8	+0,7	+0,5	+0,5	+0,4	+0,5	+0,4	+0,4	+0,2	+0,2	0,0

Scanned with OKEN Scanner

Tabela 3.3a

13	14	15	16	17	18	19	20	21	22	23	24	25
−5,3	−6,0	+2,0	−2,8	+0,2	−2,5	−4,3	−6,4	+0,1	+1,7	+1,6	+0,1	0,0
−5,6	+0,8	−2,8	+0,5	−2,6	−3,9	−5,7	+1,4	+1,7	+1,0	−0,7	−0,1	
+1,2	−2,2	+0,5	−2,3	−4,0	−5,3	+2,1	+3,0	+1,0	−1,3	−0,9		
−1,8	+1,1	−2,3	−3,7	−5,4	+2,5	+3,7	+2,3	−1,3	−1,5			
+1,5	−1,7	−3,7	−5,1	+2,4	+4,1	+3,0	0,0	−1,5				
−1,3	−3,1	−5,1	+2,7	+4,0	+3,4	+0,7	−0,2					
−2,7	−4,5	+2,7	+4,3	+3,3	+1,1	+0,5						
−4,1	+3,3	+4,3	+3,6	+1,0	+0,9							
+3,7	+4,9	+3,6	+1,3	+0,8								
+5,3	+4,2	+1,3	+1,1									
+4,6	+1,9	+1,1										
+2,3	+1,7											
+2,1												

18,5	18,9	19,5	19,5	19,8	19,7	20,1	20,8	22,1	22,1	21,5	20,7	20,6
+0,1	+0,0	+0,2	+0,0	0,0	−0,2	−0,1	−0,0	+0,1	+0,1	0,0	0,0	0,0

Scanned with OKEN Scanner

Anul	$x(t)$	1	2	3	4	5	6	7	8	9	10	11	12
1951	18,8	+3,6	+1,3	+4,5	−1,5	−0,5	−2,8	+0,5	−2,7	+0,3	−1,5	−0,9	−1,6
1952	16,4	+1,2	+4,5	−1,7	−0,5	−2,8	+0,7	−2,7	+0,4	−1,5	−0,8	−1,5	−2,2
1953	19,6	+4,4	−1,7	−0,7	−2,8	+0,7	−2,5	+0,4	−1,4	−0,8	−1,4	−2,1	−1,0
1954	13,4	−1,8	−0,7	−3,0	+0,7	−2,5	+0,6	−1,4	−0,7	−1,4	−2,0	−0,9	+3,3
1955	14,4	−0,8	−3,0	+0,5	−2,5	+0,6	−1,2	−0,7	−1,3	−2,0	−0,8	+3,4	−0,3
1956	12,1	−3,1	+0,5	−2,7	+0,6	−1,2	−0,5	−1,3	−1,9	−0,8	+3,5	−0,2	+2
1957	15,6	+0,4	−2,7	+0,4	−1,2	−0,5	−1,1	−1,9	−0,7	+3,5	−0,1	+2,6	−0,8
1958	12,4	−2,8	+0,4	−1,4	−0,5	−1,1	−1,7	−0,7	+3,6	−0,1	+2,7	−0,7	−2,77
1959	15,5	+0,3	−1,4	−0,7	−1,1	−1,7	−0,5	+3,6	0,0	+2,7	−0,6	−2,6	−4,,5
1960	13,7	−1,5	−0,7	−1,3	−1,7	−0,5	+3,8	0,0	+2,8	−0,6	−2,5	−4,6	+2,6
1961	14,4	−0,8	−1,3	−1,9	−0,5	+3,8	+0,2	+2,8	−0,5	−2,5	−4,5	+2,7	+3,7
1962	13,8	−1,4	−1,9	−0,7	+3,8	+0,2	+3,0	−0,5	−2,4	−4,5	+2,8	+3,8	+2,5
1963	13,2	−2,0	−0,7	+3,6	+0,2	+3,0	−0,3	−2,4	−4,4	+2,8	+3,9	+2,6	−0,3
1964	14,4	−0,8	+3,6	0,0	+3,0	−0,3	−2,2	−4,4	+2,9	+3,9	+2,7	−0,2	−1,1
1965	18,7	+3,5	0,0	+2,8	−0,3	−2,2	−4,2	+2,9	+4,0	+2,7	−0,1	−1,0	
1966	15,1	−0,1	+2,8	−0,5	−2,2	−4,2	+3,1	+4,0	+2,8	−0,1	−0,9		
11967	17,9	+2,7	−0,5	−2,4	−4,2	+3,1	+4,2	+2,8	0,0	−0,9			
1968	14,6	−0,6	−2,4	−4,4	+3,1	+4,2	+3,0	0,0	−0,8				
969	12,7	−2,5	−4,4	+2,9	+4,2	+3,0	+0,2	−0,8					
1970	10,7	−4,5	+2,9	+4,0	+3,0	+0,2	−0,6						
1971	18,0	+2,8	+4,0	+2,8	+0,2	−0,6							
1972	19,1	+3,9	+2,8	0,0	−0,6								
1973	17,9	+2,7	0,0	−0,8									
1974	15,1	−0,1	−0,8										
1975	14,3	−0,9											
$\bar{x}(t)$		15,2	15,1	15,1	14,9	14,9	14,9	15,1	15,1	15,2	15,2	15,3	15,4
$R_{xx}(t)$		1,00	+0,1	0,0	−0,5	−0,1	−0,2	+0,3	+0,3	+0,1	−0,4	−0,2	−0,4

Scanned with OKEN Scanner

Tabela 3.8b

13	14	15	16	17	18	19	20	21	22	23	24	25
−2,3	−1,3	+2,9	−0,4	+2,3	−0,7	−2,7	−5,2	+1,1	+2,5	+2,1	+0,4	0,0
−1,1	+0,3	−0,7	+2,4	−1,0	−2,6	−4,7	+2,1	+2,2	+1,3	−0,7	−0,4	
+3,2	−0,6	+2,1	−0,9	−2,9	−4,6	+2,6	+3,2	+1,0	−1,5	−1,5		
−0,4	+2,2	−1,2	−2,8	−4,9	+2,7	+3,7	+3,0	−1,8	−2,3			
+2,4	−1,1	−3,1	−4,8	+2,4	+3,8	+2,5	−0,8	−2,6				
−0,9	+3,0	−5,1	+2,5	+3,5	+2,6	−0,3	−1,6					
−2,8	−5,0	+2,2	+3,6	+2,3	−0,2	−1,1						
−4,8	+2,3	+3,3	+2,4	−0,5	−1,0							
+2,5	+3,4	+2,1	−0,4	−1,3								
+3,6	+2,2	−0,7	−1,2									
+2,4	−0,6	−1,5										
−0,4	−1,4											
−1,2												

15,5	15,7	15,8	15,5	15,6	15,3	15,4	15,9	16,9	16,6	15,8	14,7	14,3
+0,2	−0,4	+0,5	−0,1	−0,5	−0,1	−0,3	0,0	+0,6	+0,4	0,0	+0,1	0,0

Scanned with OKEN Scanner

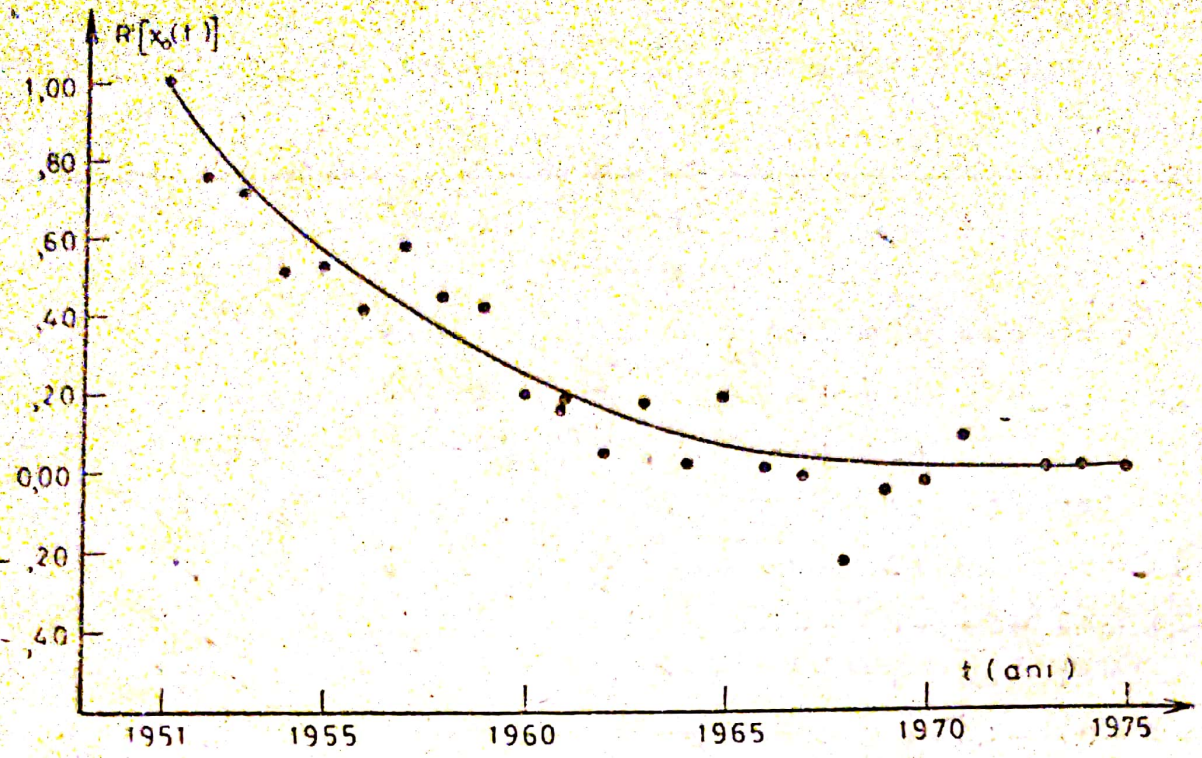


Fig. 3.8. Funcția de autocorelație normată pentru dinamica producției medii de grîu și secară la hectar în România între anii 1951 și 1975.

ține seamă, inundații sau alte cauze naturale. Chiar și anumite interese economice, neprevăzute încă într-o anumită măsură, puteau să modifice limitele de mai sus.

Prin aceste considerații am răspuns la prima din cele două întrebări pe care le-am pus mai înainte.

În penultima linie a tabelei 3.8a se găsesc valorile medii calculate pentru fiecare serie de date experimentale. Astfel, la coloana 1 pentru calcularea mediei au fost considerate toate datele din coloana a 2-a. În coloana a 4-a (marcată prin cifra 2), pentru calcularea mărimii medii nu a fost considerată mărimea 18,8, în coloana a 5-a (marcată prin cifra 3) nu am considerat mărimile 18,8 și 16,4. Pentru coloana a 6-a nu au fost considerate mărimile 18,8 16,4 și 19,6 etc. În coloana ultimă (marcată prin cifra 25) ca valoare medie a fost considerată numai ultima mărime din coloana a 2-a (14,3). Aceste mărimi medii nu reprezintă mărimea medie a producției din anul respectiv ci sînt mărimi medii ale anumitor grupuri de elemente iar variația considerabilă a acestora ne arată că nu este îndeplinită condiția de ergodicitate așa cum a fost enunțată ea mai înainte.

Pentru a răspunde la cea de-a doua întrebare care se referă la ipoteza unei variații periodice a producției de grîu și de secară la hectar, am eliminat trendul liniar, folosind în acest scop relația

$$x(t) = x_0(t) + 0{,}5218(1963 - t)$$

Scanned with OKEN Scanner

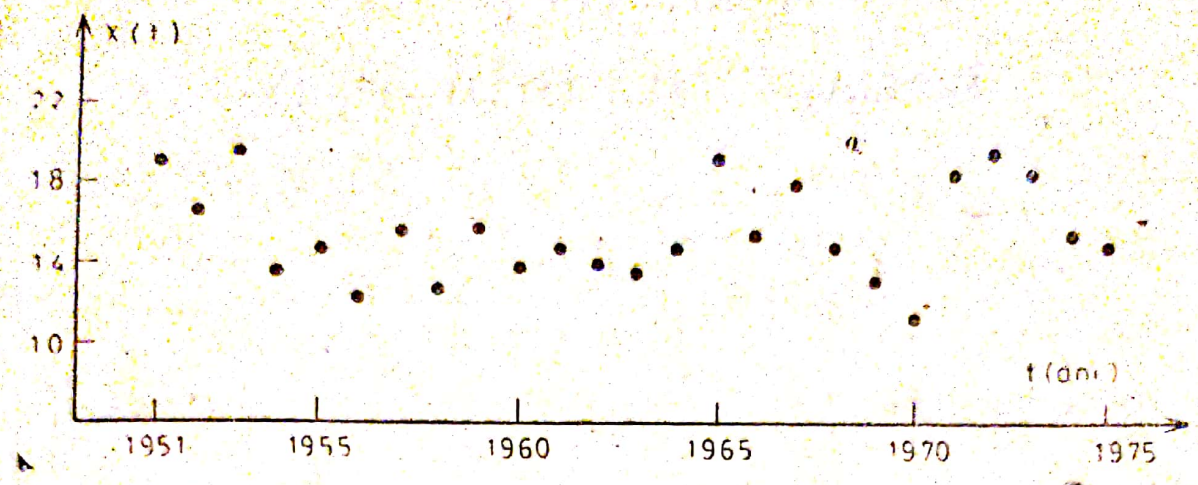


Fig. 3.9. Graficul dinamicii producției medii de grîu și secară la hectar, între anii 1951 și 1975, după eliminarea trendului liniar.

unde t este exprimat în ani și satisface inegalitățile

$$1951 \leqslant t \leqslant 1975,$$

iar $x_0(t)$ ține locul mărimilor din coloana a 2-a din tabela 3.8a. Urmînd procedeul pe care l-am descris mai sus, am întocmit tabela 3.8b, iar reprezentarea grafică a lui $x(t)$ în funcție de timp este dată pe figura 3.9 de unde se vede că este greu de susținut prezența unor fluctuații periodice. Lărgimea benzii din figura 3.9 poate fi atribuită unor variații accidentale care indică doar limitele între care poate să varieze producția de grîu și secară la hectar. Această concluzie este confirmată și de funcția de autocorelație normată care are valorile calculate și reproduse în ultima linie a tabelei 3.8b și care are serioase fluctuații accidentale, fluctuații care apar și din cauza lipsei unui număr mai mare de date numerice și care să fie repartizate pe mai mulți ani. În modul acesta, analizînd fluctuațiile funcției de autocorelație, răspundem și la cea de-a doua întrebare pe care am pus-o mai sus: oscilațiile „observate" în dinamica producției de grîu și secară nu prezintă un caracter periodic, ele sînt de natură accidentală și au un rol negativ în problemele de prognoză.

Noi am prezentat aici o aplicație practică pentru cazul unei serii mici de date, în acest mod putînd fi cu mai multă ușurință analizată și înțeleasă exemplificarea respectivă. Dar în cazurile cînd avem de elaborat o prognoză, cum era și cea de mai sus, trebuie să dispunem de date mult mai numeroase, să utilizăm un calculator electronic și astfel, într-un interval scurt de timp să putem dispune de o prognoză cît mai bună.

Observație: În cadrul aplicațiilor practice, ultimele cifre pot fi afectate de erorile de rotunjire.

CAPITOLUL IV

CORELAȚIE—CAUZALITATE—PROGNOZĂ

În capitolul precedent ne-am ocupat cu stabilirea corelației dintre diferite serii statistice și am văzut că între mărimile x_k și y_k, de exemplu, există o dependență funcțională liniară atunci cînd coeficientul de corelație are mărimea egală cu unitatea. De asemenea, pentru existența dependențelor curbilinii trebuie ca raportul de corelație să se apropie de unitate. Prin urmare, teoria corelației ne pune în evidență existența posibilă a unor relații între diferite date obținute pe cale experimentală și, prin aceasta, ne poate conduce la legătura posibilă dintre o cauză și efectul corespunzător. De multe ori, stabilirea unor corelații statistice scoate în evidență prezența anumitor relații cauzale prin intermediul cărora se va cunoaște mai bine fenomenul examinat și, în felul acesta, se va putea stabili o prognoză cît mai demnă de încredere.

În aplicațiile practice se pot ivi cazuri în care să avem de-a face cu corelații puternice sau chiar dependențe funcționale între două variabile care, însă, aparent să nu fie relații de cauzalitate. Astfel, de exemplu, între notele pe care le obțin studenții de la fizică și cei de la matematică există o corelație destul de bună. La ambele facultăți numărul studenților descrește atunci cînd considerăm notele de la 5 spre 10.

Este adevărat că între cele două serii de note, obținute la fizică și matematică, nu există o înrîurire reciprocă. O serie de note nu este cauza celeilalte serii, totuși putem remarca faptul că ele au o caracteristică comună — notele mai mari se obțin mai greu.

Prin urmare, în cadrul cercetărilor unor fenomene asemănătoare cu cel expus mai sus, se pot pune în evidență caracteristicile comune ale unor grupe de obiecte sau fenomene. De aici rezultă, apoi, concluzii privind corelațiile de grup. Astfel, dacă se examinează corelațiile dintre masele stelelor dintr-un roi și pozițiile corespunzătoare pe diagrama Hertzsprung-Russell (spectru — luminozitate) se poate stabili vîrsta roiului considerat. În acest exemplu cauzele sînt reprezentate prin condițiile fizice care au determinat formarea stelelor de diferite mase și procesul de evoluție stelară. O stea evoluează cu atît mai repede (avansează pe diagrama H—R) cu cît aceasta are masa mai mare.

Scanned with OKEN Scanner

Aşa cum am văzut mai înainte, corelaţia statistică constă în determinarea legăturii comune dintre două sau mai multe caracteristici, legătură care poate avea una sau mai multe cauze comune sau să fie pur şi simplu întîmplătoare.

În cadrul unei corelaţii statistice se pune în evidenţă legătura comună — generală, pe cînd într-o relaţie de cauzalitate se evidenţiază dependenţa dintre variabilele respective; dintre acestea cel puţin una este cauza unui efect care se simte în modul de comportare a celorlalte variabile.

Din punctul de vedere al succesiunii cunoaşterii, corelaţiile statistice pot fi indicatori ai existenţei relaţiilor cauzale însă nu întotdeauna cînd se constată o corelaţie va exista şi o cauzalitate directă. De aceea, termenul de *cauzalitate* va trebui să fie legat de prezenţa corelaţiilor prin noţiunea de *probabilitate*, adică la început corelaţia indică prezenţa anumitor cauzalităţi probabile, cauzalităţi a căror existenţă trebuie confirmată şi prin alte metode de cercetare.

Dacă am suspectat sau determinat anumite relaţii cauzale, metodele corelaţiei statistice ne permit să confirmăm sau să infirmăm anumite ipoteze făcute. Cu alte cuvinte, corelaţiile statistice constituie proba relaţiilor cauzale, iar discuţiile asupra relaţiilor cauzale sînt oportune doar în prezenţa unor corelaţii statistice demonstrate. Prin urmare, între relaţiile cauzale şi corelaţiile statistice există o strictă interdependenţă a cărei descoperire este de mare importanţă pentru problemele de prognoză, mai ales atunci cînd dorim ca la baza acesteia din urmă să aşezăm anumite modele matematice. De aceea, în capitolul prezent vom încerca să stabilim relaţii matematice — formule empirice pentru două sau mai multe serii de date experimentale şi vom prezenta unele metode pentru examinarea relaţiilor funcţionale în cazurile cînd $r \approx 1$ sau $\eta \approx 1$.

4.1. Stabilirea formulelor empirice

În urma unor activităţi practice, deseori se ajunge la un volum mare de date numerice în cadrul cărora pentru o anumită valoare x_k s-a obţinut mărimea corespunzătoare y_k. În astfel de cazuri datele înregistrate se trec în tabele care, uneori, sînt destul de voluminoase. Iată de ce, dacă din analiza matematică a datelor disponibile se stabileşte o formulă matematică — o formulă empirică — am putea fi scutiţi de a mai reproduce astfel de tabele. O formulă empirică, prin intermediul unor constante, pentru orice x_k ne-ar furniza mărimea corespunzătoare y_k. În plus, stabilirea unor formule de acest fel ne-ar putea da unele informaţii cu privire la natura fenomenului analizat sau legătura posibilă dintre acesta şi lumea înconjurătoare.

Scanned with OKEN Scanner

Pe de altă parte, stabilirea unor formule empirice ne conduce mai repede la stabilirea legăturii dintre cauză și efect, cauza sau cauzele principale care stau la baza modului de manifestare a fenomenului studiat și, prin aceasta, ajungem la o extrapolare mai corectă și o prognoză mult mai demnă de încredere.

În anumite cazuri putem cunoaște natura generală a fenomenului studiat: pulsația unei stele, variația strălucirii unor stele variabile, creșterea nivelului apei unui rîu în funcție de cantitatea de apă căzută, creșterea producției de cereale la hectar în funcție de cantitatea de îngrășămînt aplicată pe metru pătrat etc. În astfel de situații putem dispune de formule matematice — formule analitice — în care dependența lui y de x este exprimată prin intermediul anumitor constante pe care trebuie să le determinăm cu ajutorul datelor experimentale. Rezultatele astfel obținute se concretizează în așa-numitele *formule semi-empirice*.

In cadrul expunerii pe care o urmărim aici vom considera cazul general în care un anumit fenomen este observat și analizat. Aici poate fi vorba de erupția unui vulcan, izbucnirea unui cutremur de Pămînt, apariția unei furtuni, a unei epidemii, intensificarea activității atmosferice etc, iar modul de stabilire a formulelor empirice îl vom prezenta pentru diferite situații, după cum poate să se prezinte aspectul dependenței funcționale dintre seriile de date experimentale. Aplicarea acestor formule într-un caz concret, la un anumit domeniu de cercetare rămîne pe seama specialistului care a efectuat înregistrarea datelor experimentale. Cu alte cuvinte, noi vom discuta aici unele dintre cele mai frecvente situații ce se pot întîlni, vom discuta modul în care se pot stabili formule empirice și metodica prin care astfel de relații se pot utiliza pentru probleme de prognoză.

Să considerăm că, prin metodele corelației statistice, am ajuns la concluzia că o serie de date experimentale definesc o funcție de forma

$$y = f(x). \tag{4.1}$$

Dacă nu cunoaștem forma analitică a funcției (4.1) sau dacă această formă ar fi mult prea complicată și greu de utilizat practic, se pune problema ca, folosind datele experimentale — funcția empirică tabulată, să stabilim o formulă matematică, o formulă empirică

$$\tilde{y} = \tilde{f}(x) \tag{4.2}$$

astfel încît pentru $x = x_i$, $\tilde{y}_i$ să fie cît mai puțin posibil diferit de y_i. O astfel de formulă trebuie să fie cît mai simplă și ușor de utilizat în problemele de interpolare, extrapolare și prognoză. Prin urmare, considerînd că dispunem de o funcție empirică dată sub formă de tabelă, ne propunem s-o prezentăm sub o formă matematică care să

Scanned with OKEN Scanner

fie ușor de manevrat. De această problemă ne vom ocupa în prezentul capitol.

Așa cum am mai amintit și în capitolul precedent, trebuie să avem în vedere faptul că în natură pot să existe mai multe cauze care să determine un anumit efect. Nu întotdeauna dependențele funcționale au loc numai între variabilele x și y, dependența dintre acestea poate să mai depindă și de alți factori, pot să existe corelații multiple. De aceea, problema dependenței funcționale dintre x și y se pune abia atunci cînd se poate neglija influența celorlalte mărimi care ar mai putea acționa asupra acestor două variabile, sau cînd se poate admite că toate celelalte argumente păstrează valori constante (sau aproape constante) în tot timpul în care au fost adunate datele experimentale.

De fapt, din analiza unei serii de date experimentale x_k, y_k se ajunge la concluzia că acestea ar putea fi reprezentate de o clasă destul de mică de funcții: funcții liniare, funcția putere, polinoame de un anumit ordin etc. Dar prin utilizarea datelor experimentale, metodele de stabilire a formulelor empirice vor alege pe cea mai probabilă funcție dintr-un anumit grup. Iată de ce, prin metodele pe care le vom expune aici se va urmări determinarea celor mai probabile valori ale constantelor. Va fi aleasă acea funcție care va reprezenta cel mai bine datele empirice.

In unele cazuri, clasa C de funcții poate fi indicată de modul de distribuție a datelor experimentale, în alte situații însă această clasă de funcții poate fi sugerată de anumite considerații teoretice, deduse dintr-o cunoaștere preliminară a naturii fenomenului analizat. În ambele situații prin stabilirea clasei de funcții C și determinarea parametrilor acesteia pentru cazul dat, prognoza desfășurării lui este mult ușurată. Utilizarea unei astfel de funcții pentru un anumit x din afara intervalului de date disponibile, ne va permite să determinăm o valoare *probabilă* a ordonatei corespunzătoare și, prin aceasta, se poate elabora o prognoză „bună" cu privire la mărimea corespunzătoare y, să cunoaștem tăria sau intensitatea cu care acest fenomen se va manifesta în viitor.

Formulele empirice obținute pe baza rezultatelor experimentale, trebuie să satisfacă o serie de condiții: ele trebuie să fie destul de sigure și să asigure o prognoză cît de cît demnă de încredere, să aibă formă simplă pentru a putea fi aplicate cu ușurință și să se obțină prognoza într-un timp cît mai scurt posibil, iar structura lor să fie în concordanță cu existența fizică a fenomenului respectiv. Această din urmă condiție nu este totdeauna luată în considerare și de aceea anumite formule empirice își pierd repede valabilitatea, iar prognoza efectuată pe baza lor nu este demnă de încredere. Formulele empirice care nu țin seamă și de natura fenomenului pot furniza o prognoză compromițătoare.

Scanned with OKEN Scanner

Formulele empirice nu explică întotdeauna fenomenul cercetat, ele sînt mai mult niște ipoteze pe care trebuie să le verificăm prin noi serii de activități experimentale și să le supunem la noi considerații teoretice. Cu toate acestea, importanța lor este foarte mare deoarece ele duc la importante descoperiri științifice tocmai prin faptul că pot indica anumite trăsături ale fenomenului considerat.

În practică, la stabilirea unor formule empirice trebuie să avem în vedere următoarele două etape mai importante.

i) Trebuie aleasă forma generală a expresiei analitice cu care dorim să aproximăm funcția empirică disponibilă sub formă de tabelă. Pînă în prezent, însă, nu dispunem de o regulă generală în baza căreia să alegem cea mai potrivită formă a unei astfel de expresii matematice. De aceea se consideră că alegerea celei mai probabile forme, care să fie utilizată cu ușurință în problemele de prognoză, depinde de experiența și iscusința cercetătorului.

O orientare generală pentru alegerea formei expresiei matematice, cu care dorim să reprezentăm anumite date experimentale, poate fi furnizată de reprezentarea grafică. În acest scop se poate utiliza sistemul de coordonate cartezian sau alt sistem special de coordonate (semi-logaritmice, logaritmice etc) și, eventual, graficul astfel obținut să-l comparăm cu alte grafice ale căror expresii matematice sînt cunoscute. La construirea unor astfel de grafice trebuie să avem grijă ca, pe cele două axe de coordonate, să alegem scale potrivite, iar graficul să ne furnizeze maximul posibil de informație, adică graficul să nu fie deformat printr-o alegere nepotrivită a unităților.

ii) Trebuie determinate cele mai probabile valori ale constantelor care intervin în expresiile alese. Formulele trebuie să reprezinte cît mai bine posibil datele experimentale disponibile, fapt care să ne dea o mai mare încredere în problemele de extrapolare și utilizarea lor în prognoză.

În general, determinarea celor mai probabile valori ale parametrilor care definesc o funcție empirică nu prezintă dificultăți. Există totuși anumite reguli pe care trebuie să le respectăm. Trebuie să știm să alegem cea mai potrivită metodă pentru fiecare caz concret, pentru fiecare fel de date experimentale și să putem evalua limitele preciziei la care ne putem aștepta.

În cazul funcțiilor empirice, y_i poate fi afectat de eroarea $\pm\delta y_i$ iar x_i să fie afectat de eroarea $\pm\delta x_i$. Într-o situație ca aceasta, punctul $M_i(x_i, y_i)$ se poate găsi în interiorul dreptunghiului care are laturile $2\delta y_i$ și $2\delta x_i$. De aceea, atunci cînd ducem curba cu mîna liberă printre punctele respective (puncte observate) va fi suficient ca aceasta să treacă prin interiorul dreptunghiurilor erorilor respective. Dimensiunile dreptunghiului erorilor pot fi apreciate și după lărgimea benzii în interiorul căreia se află punctele observate, bineînțeles dacă aceste puncte sînt destul de dese.

Scanned with OKEN Scanner

Din cercetarea experimentală a unor fenomene sau obiecte, s-ar putea să rezulte relaţii care, aparent, să se prezinte asemănător cu una din formele (2.3), (2.4), (2.5) sau alte relaţii care se pot aduce la aceste forme. De aceea, pentru a stabili forma reală a relaţiei dintre mărimile înregistrate x_k, y_k sînt necesare şi consideraţii teoretice asupra subiectului în discuţie. Astfel, de exemplu, în anul 1911, J. Halm, utilizînd mai multe serii de date observaţionale, a cercetat corelaţiile masă-spectru şi spectru-luminozitate la un anumit grup de stele. În acest fel, el ajunge la concluzia că există o relaţie masă — luminozitate care este deosebit de importantă pentru teoria structurii interne a stelelor.

Relaţia masă-luminozitate a fost fundamentată teoretic în 1917 de către A. Eddington, iar în 1919 E. Hertzsprung stabileşte următoarea formulă empirică

$$\log \mathfrak{M} = -0{,}06(M_v - 5)$$

unde $\mathfrak{M}$ este masa stelei iar M_v reprezintă magnitudinea absolută vizuală a stelei corespunzătoare.

Dacă notăm: $\log \mathfrak{M} = y$, $M_v = x$, formula lui Hertzsprung devine

$$y = -0{,}06 \cdot x + 0{,}30$$

care este o formă liniară simplă.

Dar, dacă se are în vedere problema astrofizică despre care este vorba aici, scriem

$$\mathfrak{M} = 10^{-0{,}06(M_v - 5)},$$

formă care era imposibil de ghicit numai pe baza datelor experimentale.

În general o expresie de forma

$$y = a \cdot x^n \tag{4.3}$$

se poate scrie

$$\log y = \log a + n \cdot \log x$$

sau

$$Y = A + n \cdot X$$

unde am notat: $\log y = Y$, $\log x = X$, $\log a = A$.

Exemple de astfel de relaţii se pot aduce multe, dar utilizarea lor în probleme de prognoză este dificilă.

În cazul cînd folosim o formulă sub formă de polinom

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m \qquad (2.5) = (4.4)$$

ultimul termen care trebuie reţinut va fi indicat de teorema

Scanned with OKEN Scanner

Dacă valorile lui x sînt în progresie aritmetică iar diferențele de ordinul m sînt constante, ultimul termen necesar este acela care conține pe x^m.

Menționăm că este vorba aici de diferențele dintre ordonatele y ale punctelor succesiv observate sau determinate.

Să considerăm că, din activitatea practică, dispunem de seria de date empirice

$$x_1,\ x_2,\ \ldots,\ x_s$$
$$y_1,\ y_2,\ \ldots,\ y_s$$
$$n_1,\ n_2,\ \ldots,\ n_s$$

unde n este numărul de apariții ale perechii $(x_i,\ y_i)$ în tabela experimentală.

Pentru a exemplifica modul în care, cu aceste date empirice, putem stabili o formulă empirică, vom presupune că între x_h și y_h există o relație de forma (4.4), unde $a_0,\ a_1.\ \ldots,\ a_m$ sînt constante și se pot determina prin metoda celor mai mici pătrate, aceasta din urmă putînd fi prezentată sub aspecte, aparent, diferite.

4.2. Utilizarea metodei celor mai mici pătrate în forma elaborată de Gauss

În cazul cînd $n_h \neq 1$, pentru determinarea coeficienților a_i, în cadrul metodei celor mai mici pătrate, avem condiția:

$$S = \sum_{h=1}^{s} n_h(y - a_0 - a_1x - a_2x^2 - a_3x^3 - \ldots - a_mx^m)_h^2 = \text{minim} \qquad (4.5)$$

unde $h = \overline{1,s}$. Prin urmare,

$$\frac{\partial S}{\partial a_0} = \frac{\partial S}{\partial a_1} = \frac{\partial S}{\partial a_2} = \ldots = \frac{\partial S}{\partial a_m} = 0 \qquad (4.5a)$$

Efectuînd operațiile din (4.5) și (4.5a) rezultă următorul sistem de ecuații normale:

$$\left.\begin{aligned}
&a_0\Sigma n_h + a_1\Sigma n_h x_h + a_2\Sigma n_h x_h^2 + \ldots + a_m\Sigma n_h x_h^m = \Sigma n_h y_h\\
&a_0\Sigma n_h x_h + a_1\Sigma n_h x_h^2 + a_2\Sigma n_h x_h^3 + \ldots + a_m\Sigma n_h x_h^{m+1} = \Sigma n_h y_h x_h\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&a_0\Sigma n_h x_h^m + a_1\Sigma n_h x_h^{m+1} + a_2\Sigma n_h x_h^{m+2} + \ldots + a_m\Sigma n_h x_h^{2m} = \Sigma n_h y_h x_h^m
\end{aligned}\right\} \qquad (4.6)$$

Rezolvarea sistemului (4.6) se bazează pe faptul că din datele experimentale dispunem de mărimile x_h și y_h cu numerele corespun-

Scanned with OKEN Scanner

zătoare n_h, astfel că singurele necunoscute sînt constantele a_i unde $i = \overline{1, m}$.

Pentru rezolvarea unor sisteme de ecuații normale de forma (4.6) există diferite scheme de calcul, autorul a folosit schema lui Gauss (vezi Todoran, 1976, cap. II, p. 85—86). Marinescu și colaboratorii (I.Marinescu et al, 1969) recomandă metoda simplex. În cadrul prezentei expuneri noi am folosit un procedeu relativ mai simplu care se bazează pe eliminarea succesivă a necunoscutelor. În acest scop în (4.6) împărțim fiecare ecuație cu coeficientul necunoscutei a_0 și avem mai departe.

$$\left.\begin{aligned}
&a_0 + a_1\frac{\Sigma n_h x_h}{\Sigma n_h} + a_2\frac{\Sigma n_h x_h^2}{\Sigma n_h} + \ldots + a_m\frac{\Sigma n_h x_h^m}{\Sigma n_h} = \frac{\Sigma n_h y_h}{\Sigma n_h}\\
&a_0 + a_1\frac{\Sigma n_h x_h^2}{\Sigma x_h x_h} + a_2\frac{\Sigma n_h x_h^3}{\Sigma n_h x_h} + \ldots + a_m\frac{\Sigma n_h x_h^{m+1}}{\Sigma n_h x_h} = \frac{\Sigma n_h y_h x_h}{\Sigma n_h x_h}\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&a_0 + a_1\frac{\Sigma n_h x_h^{m+1}}{\Sigma n_h x_h^m} + a_2\frac{\Sigma n_h x_h^{m+2}}{\Sigma n_h x_h^m} + \ldots + a_m\frac{\Sigma n_h x_h^{2m}}{\Sigma n_h x_h^m} = \frac{\Sigma n_h y_h x_h^m}{\Sigma n_h x_h^m}
\end{aligned}\right\} \quad (4.7)$$

Apoi, din prima ecuație scădem toate celelalte ecuații. În modul acesta se elimină o necunoscută iar numărul de ecuații va avea cu o ecuație mai puțin. Continuînd în felul acesta vom ajunge la sisteme cu ecuații și necunoscute din ce în ce mai puține. Astfel, antepenultimul sistem va avea trei ecuații cu trei necunoscute, penultimul sistem va avea două ecuații cu două necunoscute și, în final, o ecuație cu o necunoscută care se determină imediat.

Prin substituțiile corespunzătoare se determină celelalte necunoscute.

Unul dintre neajunsurile aproximării unei curbe empirice cu un polinom de gradul m este și faptul că nu știm dinainte care este cea mai potrivită valoare pentru gradul respectiv, mai ales cînd avem de-a face cu un număr mare de date experimentale. De aceea, în practică, uneori se fac încercări și se rezolvă sistemul (4.5) pentru diferite valori ale lui m, iar valoarea acceptată va fi aceea pentru care suma (4.5) rămîne minimă și nu se reduce cînd crește m. Prin urmare, în (4.6) x_h și y_h sînt mărimi cunoscute din cadrul seriei de date empirice, iar a_i se determină prin calcule. Aici sînt necesare foarte multe calcule pînă cînd se determină valoarea cea mai probabilă a lui m. În plus, de fiecare dată cînd se schimbă valoarea lui m, trebuie să refacem toate calculele de la început și pînă la sfîrșit.

Pentru evitarea unor neajunsuri de felul celor menționate mai sus, adesea este convenabil să ne folosim de polinoamele ortogonale, asupra utilizării acestora ne vom opri în paragraful următor.

Scanned with OKEN Scanner

4.3. Metoda polinoamelor ortogonale

Să considerăm polinomul (4.4) scris sub forma:

$$y = A_0 p_0(x) + A_1 p_1(x) + \ldots + A_m p_m(x) \qquad (4.8)$$

în care $p_i(x)$ înseamnă polinoame în x de ordinul i $(i = \overline{1, m})$ și care satisfac condiția de ortogonalitate,

$$\sum_{h=1}^{s} n_h p_i(x_h) p_k(x_h) = (p_i p_k) = 0, \qquad (i \neq k) \qquad (4.9)$$

$$\sum_{h=1}^{s} n_h p_i(x_h) p_i(x_h) = (p_i p_i) \neq 0, \qquad (i = k) \qquad (4.10)$$

unde avem $k = \overline{0, m}$.

Utilizînd polinomul (4.8), suma (4.5) devine:

$$S = \sum_{h=1}^{s} n_h \left[y_h - \sum_{i=0}^{m} A_i p_i(x_h) \right]^2 \qquad (4.11)$$

iar coeficienții A_i se determină din condițiile de minim ale acestei sume, adică:

$$\frac{\partial S}{\partial A_k} = 0, \qquad (k = \overline{0,m}) \qquad (4.12)$$

Din (4.11) și (4.12) avem imediat:

$$\frac{\partial S}{\partial A_k} = 2 \sum_{h=1}^{s} n_h \left[y_h - \sum_{i=0}^{m} A_i p_i(x_h) \right] p_k(x_h) = 0$$

sau

$$\sum_{h=1}^{s} n_h p_k(x_h) \sum_{i=0}^{m} A_i p_i(x_h) = \sum_{h=1}^{s} n_h y_h p_k(x_h)$$

care se mai poate scrie:

$$\sum_{i=0}^{m} A_i \sum_{h=1}^{s} n_h p_i(x_h) p_k(x_h) = \sum_{h=1}^{s} n_h y_h p_k(x_h) \qquad (4.13)$$

În (4.13) se ține seamă de (4.9) și (4.10), așa că avem mai departe:

$$\sum_{i=0}^{m} A_i (p_i p_k) = (y p_k) \qquad (k = \overline{0,m})$$

ceea ce revine la a reține numai termenii în care $i = k$, adică:

$$A_k (p_k p_k) = (y p_k) \qquad (4.14)$$

Scanned with OKEN Scanner

de unde rezultă imediat :

$$A_k = \frac{(yp_k)}{(p_k p_k)} \tag{4.15}$$

Prin urmare polinomul (4.8) se poate scrie :

$$y = \sum_{k=0}^{m} \frac{(yp_k)}{(p_k p_k)} p_k(x) \tag{4.16}$$

și care îndeplinește condiția de minim a sumei (4.11).

Din (4.15) se vede că determinările coeficienților A_k sînt independente între ele și pot fi în număr oricît de mare, atunci cînd dispunem de un număr destul de mare de polinoame ortogonale p_i. În acest mod ne putem opri la un A_K pentru care devin nesemnificative calculele pentru coeficienții de un rang mai mare. Astfel, dacă pentru polinomul (4.8) s-au reținut m termeni și s-au determinat coeficienții A_k $(k = \overline{0,m})$ și nu sîntem mulțumiți de valorile acestora pînă la acest ordin, nu avem decît să mai calculăm cîțiva coeficienți A_{m+1}, A_{m+2} etc, calculele precedente rămînînd valabile și pe mai departe.

În cazul cînd utilizăm soluția (4.16), suma (4.11) devine :

$$S_m = \sum_{m}^{2} = \sum_{h=1}^{s} n_h \left[y_h - \sum_{k=0}^{m} \frac{(yp_k)}{(p_k p_k)} p_k(x_h) \right]^2 \tag{4.17}$$

sau, efectuînd mai departe calculele :

$$S_m = \sum_{m}^{2} = \sum_{h=1}^{s} n_h y_h^2 + \sum_{k=0}^{m} \frac{(yp_k)^2}{(p_k p_k)^2} \sum_{h=1}^{s} n_h p_k^2(x_h) - 2 \sum_{k=0}^{m} \frac{(yp_k)}{(p_k p_k)} \sum_{h=1}^{s} n_h y_h p_k(x_h)$$

sau

$$S_m = \sum_{m}^{2} = \sum_{h=1}^{s} n_h y_h^2 + \sum_{k=0}^{m} \frac{(yp_k)^2}{(p_k p_k)} \left(\frac{\sum_{h=1}^{s} n_h p_k(x_h)\, p_k(x_h)}{(p_k p_k)} \right) -$$

$$- 2 \sum_{K=0}^{m} \frac{(yp_k)}{(p_k p_k)} \sum_{h=1}^{s} n_h y_h p_k(x_h)$$

Aici ținem seamă de notațiile (4.9) și (4.10) astfel încît :

$$S_m = \sum_{m}^{2} = (yy) + \sum_{k=0}^{m} \frac{(yp_k)^2}{(p_k p_k)} - 2 \sum_{K=0}^{m} \frac{(yp_k)^2}{(p_k p_k)}$$

adică :

$$S_m = \sum_{m}^{2} = (yy) - \sum_{k=0}^{m} \frac{(yp_k)^2}{(p_k p_k)} = (yy) - \sum_{k=0}^{m} \frac{(yp_k)}{(p_k p_k)} (yp_k) \tag{4.18}$$

Scanned with OKEN Scanner

unde ținem seamă de relația (4.15) și de faptul că $i = k$, astfel că

$$S_m = \sum_m^2 = (yy) - \sum_{k=0}^{m} A_k(yp_k) \tag{4.19}$$

În mod analog, pentru gradul $m + 1$ vom avea:

$$S_{m+1} = \sum_{m+1}^{2} = (yy) - \sum_{k=0}^{m} A_k(yp_k) - A_{m+1}(yp_{m+1}) = \sum_m^2 - A_{m+1}(yp_{m+1}) \tag{4.20}$$

de unde, ținînd seamă de (4.15), putem scrie:

$$A_{m+1} = \frac{(yp_{m+1})}{(p_{m+1}p_{m+1})} \tag{4.21}$$

astfel că (4.20) devine:

$$S_{m+1} = \sum_{m+1}^{2} = \sum_m^2 - \frac{(yp_{m+1})^2}{(p_{m+1}p_{m+1})} \tag{4.22}$$

Prin urmare, pentru a trece de la polinomul de ordinul m la un polinom de ordinul $m + 1$, nu-i necesar să refacem toate calculele din nou pentru a calcula suma $S_{m+1} = \sum_{m+1}^{2}$, trebuie numai ca din $\sum_m^2$ calculat anterior să scădem termenul suplimentar:

$$\frac{(yp_{m+1})^2}{(p_{m+1}p_{m+1})}$$

care se calculează cu totul independent de calculele efectuate anterior. În modul acesta se elimină neajunsul semnalat mai sus în legătură cu necesitatea refacerii tuturor calculelor de fiecare dată cînd schimbăm gradul polinomului respectiv.

Pînă acum am insistat asupra modului de determinare a coeficienților A_i sau A_k $(i = k)$ ai polinomului (4.8) unde, în mod tacit, am presupus cunoscute expresiile polinoamelor ortogonale $p_i(x)$. Prin urmare, chestiunea care ne-a mai rămas este legată de determinarea acestora în funcție de datele obținute pe cale experimentală.

În acest scop ne vom servi de faptul că polinoamele ortogonale $p_i(x)$ pot fi alese astfel încît întotdeauna să avem $p_0(x) = 1$. În cazul cînd în (4.8) constanta A_0 ar fi înmulțită cu un polinom diferit de unitate $(\neq 1)$, și dacă este diferit și de zero $(\neq 0)$, am putea împărți cu acel polinom, am avea astfel niște rapoarte de polinoame ortogonale care și ele, la rîndul lor, ar fi tot polinoame ortogonale. Așadar, forma generală a polinoamelor ortogonale va fi:

$$p_n(x) = x^n + a_{n-1}p_{n-1} + \ldots + a_1p_1(x) + a_0p_0(x) \tag{4.23}$$

Scanned with OKEN Scanner

Evident aici am considerat $a_n = 1$, dar dacă a_n ar fi fost diferit de unitate ($\neq 1$) și diferit de zero ($\neq 0$), am fi putut împărți cu el și am fi obținut tot polinoame ortogonale.

Dacă în (4.23) înmulțim succesiv cu $p_{n-1}, p_{n-2}, \ldots, p_0$ și după fiecare înmulțire însumăm și ținem seamă de condițiile de ortogonalitate (4.9) și (4.10), se determină expresiile coeficienților respectivi. Prin urmare, dacă în (4.23) înmulțim cu $p_{n-1}(x)$, însumăm și ținem seamă de (4.9) și (4.10), avem:

$$(p_n p_{n-1}) = 0 = (x^n p_{n-1}) + a_{n-1}(p_{n-1} p_{n-1})$$

de unde

$$a_{n-1} = -\frac{(x^n p_{n-1})}{(p_{n-1} p_{n-1})} \tag{4.24}$$

Dacă efectuăm celelalte înmulțiri, se obțin expresii asemănătoare cu (4.24). Dealtfel, expresia (4.24) poate fi considerată ca o formă generală a coeficienților din (4.24). În modul acesta, pentru poligoanele ortogonale $p_i(x)$ avem relațiile:

$$\left.\begin{aligned}
&p_0(x) = 1 \\
&p_1(x) = x - \frac{(x p_0)}{(p_0 p_0)} p_0(x) = x - \frac{\Sigma n_h x_h}{\Sigma n_h} = x - \bar{x} \\
&p_2(x) = x^2 - \frac{(x^2 p_1)}{(p_1 p_1)} p_1(x) - \frac{(x^2 p_0)}{(p_0 p_0)} p_0(x) \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&p_n(x) = x^n - \frac{(x^n p_{n-1})}{(p_{n-1} p_{n-1})} p_{n-1}(x) - \ldots - \frac{(x^n p_0)}{(p_0 p_0} p_0(x)
\end{aligned}\right\} \tag{4.25}$$

Proprietetea de ortogonalitate a polinoamelor (4.23) se verifică cu ușurință. În acest scop, scriem:

$$\begin{aligned}
&(p_0 p_0) = \Sigma n_h p_0(x) p_0(x) = \Sigma n_h \neq 0 \\
&(p_0 p_1) = \Sigma n_h p_0(x) p_1(x) = \Sigma n_h (x - \bar{x}) = \Sigma n_h x_h - \Sigma n_h \bar{x} = 0 \\
&(p_1 p_1) = \Sigma n_h p_1(x) p_1 x() = \Sigma n_h p_1^2(x) \neq 0 \\
&\text{etc.}
\end{aligned}$$

În vederea reducerii volumului de calcul, în (4.25) înmulțim prima relație cu p_0, a doua cu p_1 etc. și ținem seamă de condițiile de ortogonalitate (4.9) și (4.10). În acest caz putem scrie mai departe:

$$\left.\begin{aligned}
&p_0(x) p_0(x) = 1 \\
&p_1(x) p_1(x) = x_h p_1(x_h) \\
&\text{etc.}
\end{aligned}\right\} \tag{4.26}$$

Scanned with OKEN Scanner

$$\left.\begin{aligned}
&(p_0 p_0) = \Sigma n_h = n \\
&(p_1 p_1) = \Sigma n_h p_1(x_h) p_1(x_h) = (x p_1) = \Sigma n_h x_h p_1(x_h) \\
&\dots\dots\dots\dots\dots\dots\dots\dots \\
&(p_m p_m) = (x^m p_m) = \Sigma n_h x_h^m p_m(x_h)
\end{aligned}\right\} \qquad (4.27)$$

Utilizarea expresiilor (4.27) ne scuteşte de a mai calcula valorile numerice ale polinoamelor $p_i(x)$ şi, totodată, ne permit să calculăm cu mai multă uşurinţă sumele $(p_i p_i)$ şi respectiv $(x^i y)$ care intervin în relaţiile (4.15) pentru determinarea coeficienţilor A_k şi în relaţiile (4.25) pentru determinarea expresiilor poligoanelor ortogonale $p_i(x)$.

Aplicaţii numerice

Pentru ilustrarea modului de utilizare practică a schemelor prezentate mai sus vom da aici două exemple: unul pe care-l reluăm după lucrarea lui P.L.Cebîşev (1955) în care ilustrăm toate etapele de calcul. În al doilea vom considera un caz destul de simplu în care este scoasă în evidenţă nesiguranţa unor formule empirice. Se vor analiza precauţiunile pe care trebuie să le luăm atunci cînd este vorba de utilizarea unor formule empirice în problemă de prognoză.

a) În vederea aplicării practice a metodei celor mai mici pătrate pentru stabilirea unei formule empirice ne vom referi la secvenţa de date numerice utilizate de Cebîşev. Reproducem aceste date în primele două coloane ale tebelei 4.1. Celelalte coloane au semnificaţia specificată şi le-am întocmit în vederea utilizării lor în formulele prezentate mai sus.

Din compararea valorilor calculate în tabela 4.1 şi sistemul de ecuaţii normale (4.6) constatăm că este vorba de aproximarea datelor

Tabela 4.1

x_h	y_h	x_h^2	x_h^3	x_h^4	y_h^2	$x_h y_h$	$x_h^2 y_h$
0,15411	19,47	0,02375	0,00337	0,00056	379,08	3,00052	0,46241
0,19516	21,83	0,03809	0,00743	0,00145	476,55	4,26034	0,83145
0,22143	23,11	0,04903	0,01086	0,00240	534,07	5,11725	1,13311
0,28802	26,11	0,08295	0,02389	0,00688	681,73	7,52020	2,16596
0,32808	27,60	0,10764	0,03531	0,01158	761,76	9,05501	2,97075
0,38183	28,89	0,14579	0,05567	0,02126	834,63	11,03105	4,21199
0,45517	33,17	0,20718	0,09430	0,04292	1100,25	15,09799	6,87215
0,57012	33,38	0,32504	0,18531	0,10565	1114,22	19,03060	10,84949
0,75930	32,31	0,57654	0,43776	0,33240	1043,94	24,53298	18,62790
0,91075	31,88	0,82947	0,75544	0,68801	1016,33	29,03471	26,44337
1,13895	25,46	1,29721	1,47745	1,68275	648,21	28,99767	33,02691
5,40292	303,21	3,68269	3,08709	2,89586	8590,77	156,67832	107,59549

Scanned with OKEN Scanner

experimentale cu un polinom de gradul doi. În aceste condiții sistemul de ecuații normale devine:

$$\left.\begin{array}{rl} 11\,a_0 + 5{,}40292\,a_1 + 3{,}68269\,a_2 = 303{,}21 & 323{,}29561 \\ 5{,}40292\,a_0 + 3{,}68269\,a_1 + 3{,}08709\,a_2 = 156{,}67832 & 168{,}85102 \\ 3{,}68269\,a_0 + 3{,}08709\,a_1 + 2{,}89586\,a_2 = 107{,}59549 & 117{,}26113 \end{array}\right\} \quad (*)$$

(s)

unde s reprezintă suma tuturor coeficienților numerici de pe linia respectivă. Aici am păstrat numărul de zecimale egal cu acela din lucrarea lui Cebîșev, în vederea unei comparări a rezultatelor.

Pentru aplicarea metodei lui Gauss am utilizat schemele prezentate anterior (Todoran, 1976, pg. 81—82). În conformitate cu aceste scheme, aplicația numerică duce la următoarele tabele de calcul.

11,0000	5,4029	3,6827	303,21	323,2956
	3,6827	3,0871	156,6783	168,8511
0,491175	2,6538	1,8089	148,9290	158,7946
	1,0289	1,2782	7,7493	10,0565
		2,8959	107,5955	117,2611
0,33479		1,2330	101,5117	108,2361
		1,6629	6,0838	9,0250
1,242327		1,5880	9,6271	12,4934
		0,0749	−3,5433	−3,4684

$$a_2 = -3{,}5433 : 0{,}0749 \qquad a_2 = -47{,}307$$

Pentru necunoscutele a_1 și a_0 mai vin încă două tabele asemănătoare din care rezultă

$$a_1 = 3{,}0728 : 0{,}0464 = 66{,}223$$
$$a_0 = 8{,}1054 : 0{,}7476 = 10{,}842$$

Dacă ne referim la metoda eliminării, în conformitate cu (4.7) sistemul de ecuații normale (*) devine:

$$\begin{array}{l} a_0 + 0{,}4912\,a_1 + 0{,}3348\,a_2 = 27{,}5645 \\ a_0 + 0{,}6816\,a_1 + 0{,}5714\,a_2 = 28{,}9988 \\ a_0 + 0{,}8383\,a_1 + 0{,}7864\,a_2 = 29{,}2165 \end{array} \quad (**)$$

De aici se elimină a_0 și obținem:

$$\begin{array}{l} 0{,}3471\,a_1 + 0{,}4516\,a_2 = 1{,}6520 \\ 0{,}1567\,a_1 + 0{,}2150\,a_2 = 0{,}2177, \end{array}$$

care se poate scrie:

$$\begin{array}{l} a_1 + 1{,}3009\,a_2 = 4{,}7595 \\ a_1 + 1{,}3722\,a_2 = 1{,}3898, \end{array} \quad (***)$$

Scanned with OKEN Scanner

de unde se elimină a_1 și avem:

$$a_2 = \frac{3,3697}{-0,0713} = -47,261$$

Înlocuind pe a_2 în a doua ecuație din (***), rezultă:

$$a_1 = 1,3898 + 64,8515 = 66,24$$

Mărimea lui a_0 se determină din a treia ecuație din (**) în care am înlocuit valorile lui a_1 și a_2 astfel că:

$$a_0 = 29,2165 - 55,5290 + 37,1663$$
$$a_0 = 10,854$$

O comparație între rezultatele obținute va fi dată după ce vom aplica și metoda polinoamelor ortogonale.

b) Pentru utilizarea practică a polinoamelor ortogonale facem apel tot la tabela 4.1 unde ultima linie reprezintă sumele coloanelor respective. În acest caz avem:

$$n = 11, \quad (x) = \Sigma n_h x_h = 5,40292, \quad (x^2) = \Sigma n_h x_h^2 = 3,68269,$$
$$(x^3) = \Sigma n_h x_h^3 = 3,08709, \quad (x^4) = \Sigma n_h x_h^4 = 2,89586,$$
$$(y) = \Sigma n_h y_h = 303,21, \quad (yy) = (y^2) = \Sigma n_h y_h^2 = 8590,77,$$
$$(xy) = \Sigma n_h x_h y_h = 156,67832, \quad (x_h^2 y_h) = \Sigma n_h x_h^2 y_h = 107,59549$$

Cu ajutorul acestor mărimi trebuie să calculăm produsele $A_0 p_0$, $A_1 p_1$ etc. care sînt termeni ai polinomului (4.8.) În acest scop utilizăm prima relație din (4.27), adică

$$(p_0 p_0) = 11$$

Apoi, dacă în (4.8) înmulțim cu $n_h p_0$ și însumăm, pe baza proprietăților de ortogonalitate (4.9) și (4.10), avem:

$$(y p_0) = \Sigma n_h y_h p_0 = \Sigma n_h y_h = (y) = 303,21$$

Prin urmare, efectuînd înlocuirile în (4.15), rezultă:

$$A_0 = \frac{(y p_0)}{(p_0 p_0)} = \frac{303,21}{11} = 27,5645$$
$$A_0 p_0 = 27,5645$$

unde am ținut seamă de faptul că $p_0(x) = 1$.

Pentru calcularea sumei pătratelor $\sum_0^2$, din (4.19) putem scrie:

$$S_0 = \sum_0^2 = (yy) - A_0 (y p_0) = 8590,77 - 27,5645 \times 303,21$$
$$\sum_0^2 = 232,94.$$

În mod analog procedăm și pentru termenul al doilea $A_1 p_1$. În acest scop ne folosim de a doua relație din (4.25), adică:

$$p_1(x) = x - \bar{x} = x - 0{,}491175$$

unde ne-am folosit de:

$$\bar{x} = \frac{\Sigma n_h x_h}{\Sigma n_h} = \frac{5{,}40292}{11} = 0{,}491175$$

A doua relație din (4.27) ne dă:

$$(p_1 p_1) = \Sigma n_h x_h p_1(x_h) = \Sigma n_h x_h^2 - \bar{x}\Sigma n_h x_h = (x^2) - 0{,}491175(x)$$

$$(p_1 p_1) = 3{,}68269 - 0{,}491175 \times 5{,}40292 = 1{,}028913.$$

Apoi, ținînd seamă de expresia găsită pentru $p_1(x)$, avem

$$(y p_1) = (xy) - \bar{x}(y) = 156{,}67832 - 148{,}92902 = 7{,}7493$$

Din (3.15) avem:

$$A_1 = \frac{(y p_1)}{(p_1 p_1)} = \frac{7{,}7493}{1{,}028913} = 7{,}531538$$

astfel că

$$A_1 p_1 = 7{,}531538\, x - 7{,}531538 \times 0{,}491175$$
$$A_1 p_1 = 7{,}531538\, x - 3{,}69930$$

și

$$\sum_1^2 = \sum_0^2 - A_1(y p_1) = 232{,}94 - 7{,}531538 \times 7{,}7493 = 174{,}56$$

Din compararea lui $\sum_1^2$ cu $\sum_0^2$ se vede că există o tendință de micșorare a lui $\sum_i^{12}$ odată cu luarea în considerare a unor termeni suplimentari în polinomul (4.8). De aceea vom calcula mai departe termenul $A_2 p_2(x)$ și $\sum_2^2$.

Reluînd a treia relație din (4.25)

$$p_2(x) = x^2 - \frac{(x^2 p_0)}{(p_0 p_0)} p_0(x) - \frac{(x^2 p_1)}{(p_1 p_1)} p_1(x)$$

unde

$$(x^2 p_0) = (x^2) = 3{,}68269$$

și

$$(x^2 p_1) = \Sigma n_h x_h^2 (x_h - \bar{x}) = (x^3) - \bar{x}(x^2)$$
$$(x^2 p_1) = 3{,}08709 - 3{,}68269 \times 0{,}491175 = 1{,}27825$$

Scanned with OKEN Scanner

Prin urmare

$$p_2 = x^2 - \frac{1,27825}{1,028913} p_1(x) - \frac{3,68269}{11}$$

$$p_2 = x^2 - 1,24233\, p_1(x) - 0,33479,$$

de unde, înmulțind cu $n_h y_h$ și însumînd, avem

$$(yp_2) = (x^2y) - 1,24233(yp_1) - 0,33479\ (y)$$

$$(yp_2) = -3,54334$$

În relația (*) înmulțim cu $n_h x_h^2$ și însumăm, astfel că dacă avem în vedere și ultima relație din (4.27), putem scrie

$$(p_2p_2) = (x^4) - 1,24233(x^2p_1) - 0,33479\ (x^2)$$

$$(p_2p_2) = 2,89586 - 1,58800 - 1,23293 = 0,07493$$

Efectuînd înlocuirile în (4.15) avem

$$A_2 = \frac{(yp_2)}{(p_2p_2)} = \frac{-3,54334}{0,0749335} = -47,2864$$

Apoi

$$A_2p_2 = -47,2864[x^2 - 1,24233\ p_1(x) - 0,33479]$$

$$A_2p_2 = -47,2864\ x^2 + 58,74516\ x - 13,02311$$

Din (4.22) putem scrie

$$\sum_2^2 = \sum_1^2 - A_2(yp_2) = 174,56 - 167,55 = 7,01$$

În cazul cînd nu sîntem mulțumiți de mărimea sumei $\sum_2^2$ și dorim să obținem o cantitate și mai mică, vom lua în considerare și gradul trei al polinomului considerat. Aici, însă, luarea în considerare a termenului A_3p_3 nu aduce o îmbunătățire esențială cu privire la aproximarea datelor „observaționale" examinate. Prin urmare, ținînd seamă de rezultatele de mai sus, formula empirică dedusă pentru exemplul analizat aici va fi

$$y = A_0p_0(x) + A_1p_1(x) + A_2p_2(x)$$

care devine

$$\begin{array}{rl} y = & 27,5645 \\ & -\ 3,6993 + 7,53154\ x \\ & -\ 13,0231 + 58,74516\ x - 47,2864\ x^2 \\ \hline y = & 10,8421 + 66,27670\ x - 47,2864\ x^2 \end{array}$$

Scanned with OKEN Scanner

În vederea comparării rezultatelor obținute aici, prin diferite scheme de calcul, precum și cele determinate de Cebîșev, scriem mai jos polinoamele respective în ordinea schemelor utilizate

$y = 10{,}834 + 66{,}311\ x \quad -47{,}313\ x^2$ (schema lui Cebîșev)
$y = 10{,}842 + 66{,}223\ x \quad -47{,}307\ x^2$ (schema lui Gauss)
$y = 10{,}854 + 66{,}240\ x \quad -47{,}261\ x^2$ (schema eliminării)
$y = 10{,}842 + 66{,}277\ x \quad -47{,}286\ x^2$ (schema polinoamelor ortogonale)

Desigur, faptul că în toate schemele de calcul amintite mai sus s-a ținut seamă de principiul celor mai mici pătrate, dacă nu există erori grosolane de calcul, rezultatele trebuie să fie în concordanță.

În privința volumului de calcul, metoda eliminării, așa cum am utilizat-o aici, este mai rapidă decît schema lui Gauss. În schimb, prin aplicarea metodei lui Gauss se obține și un control asupra operațiilor numerice. Aceasta din urmă furnizează și o oarecare informație asupra preciziei cu care au fost determinați coeficienții polinomului respectiv.

Așa cum am mai subliniat, utilizarea polinoamelor ortogonale se bucură de avantajul că se poate determina cu mai mare siguranță gradul polinomului prin care au fost aproximate datele experimentale.

Noi am prezentat aici o aplicație practică a mai multor scheme de calcul la aceleași date numerice tocmai pentru a scoate în evidență avantajele și dezavantajele fiecărei scheme, cercetătorul care dorește să le utilizeze putînd să decidă asupra celei mai potrivite scheme pentru o anumită situație dată.

b) În cel de al doilea exemplu dorim să scoatem în evidență pericolul la care ne putem expune atunci cînd utilizăm o formulă empirică care a fost dedusă fără a ține seamă de toate condițiile cerute de o astfel de problemă. În acest scop ne folosim de formula

$$y = x^2$$

unde pentru anumite valori ale lui x obținem mărimile corespunzătoare ale lui y. Rezultatele astfel obținute le prezentăm în primele două coloane ale tabelei 4.2.

Tabela 4.2

x	$y = x^2$	$xy = x^3$	$y^2 = x^4$	O—C
10,00	100,0000	1000,0000	10000,0000	+0,0010
10,01	100,2001	1003,0030	10040,0600	+ ,0002
10,02	100,4004	1006,0120	10080,2403	− ,0004
10,03	100,6009	1009,0270	10120,5411	− ,0008
10,04	100,8016	1012,0481	10160,9626	− ,0010
10,05	101,0025	1015,0751	10201,5050	− ,0010
10,06	101,2036	1018,1082	10242,1687	− ,0009
10,07	101,4049	1021,1473	10282,9537	− ,0005
10,08	101,6064	1024,1925	10323,8605	+ ,0001
10,09	101,8081	1027,2437	10364,8892	+ ,0009
10,10	102,0100	1030,3010	10406,0401	+0,0019
110,55	1111,0385	11166,1579	112223,2212	−0,0005

Scanned with OKEN Scanner

Din ultima linie a tabelei 4.2 avem imediat

$$n = 11.\ (x) = 110{,}55, \quad \bar{x} = 10{,}05,\ (x^2) = 1111{,}0385.$$
$$(x^3) = 11166{,}1579 = (xy), \quad (yy) = (x^4) = 112223{,}2212.$$

Cu ajutorul acestor mărimi calculăm succesiv termenii polinomului (4.8).

I. $(p_0p_0) = 11,\ (yp_0) = 1111{,}0385,\ A_0 = 101{,}0035,\ A_0p_0 = 101{,}0035$

$$S_0 = (yy) - A_0(yp_0) = 4{,}4441 = \sum_0^2$$

II. $p_1(x) = x - 10{,}05, \qquad (p_1p_1) = (x^2) - 10{,}05(x) = 0{,}0110$

$$(yp_1) = (x^3) - \bar{x}(x^2) = 0{,}2210$$

$$A_1 = 20{,}0909, \qquad A_1p_1(x) = 20{,}0909x - 201{,}9135$$

$$S_1 = \sum_1^{21} = \sum_0^{12} - A_1(yp_1) = 0{,}0040$$

Ţinînd seamă de cei doi termeni calculaţi mai sus, polinomul (4.8) se poate scrie

$$y = A_0p_0 + A_1p_1 = 20{,}0909\ x - 100{,}91 \qquad (**)$$

Cu această expresie calculăm ordonatele y_c pentru abscisele din prima coloană a tabelei 4.2 astfel ca $O-C = y - y_c$ din coloana a 5-a să ne arate modul în care funcţia (**) aproximează datele din coloana a doua din tabelă. Prin urmare, din comportarea diferenţelor O—C şi mai ales a sumei acestora, s-ar părea că funcţia (**) reprezintă o aproximaţie destul de bună pe intervalul utilizat.

Examinînd mai atent situaţia discutată aici, se vede că, în cadrul problemelor de prognoză, aproximaţia de mai sus este cu totul necorespunzătoare deoarece, aşa cum se vede şi pe graficul din figura 4.1. dreapta $y = a \cdot x + b$ aproximează arcul de parabolă pe intervalul utilizat, dar aici există deviaţii pozitive la extremităţi şi deviaţii negative la mijlocul intervalului de abscise utilizate. De aceea suma $\Sigma(O-C) = -0{,}0005$ apare a fi mică, însă atunci cînd ne vom referi la problemele de prognoză va trebui să considerăm abscise din exteriorul intervalului utilizat şi mai mari ca cele utilizate în determinarea funcţiei (**), fapt care ne conduce la o extrapolare şi, prin aceasta, va fi o prognoză care se abate din ce în ce mai mult de la starea reală a fenomenului respectiv.

Dacă urmărim desenul din figura 4.1. observăm că pe măsură ce trece timpul, pe măsură ce x creşte, linia dreaptă $(y = a \cdot x + b)$ se abate din ce în ce mai mult de la arcul de parabolă $(y = x^2)$ astfel că o prognoză bazată pe extrapolarea liniară va fi sistematic afectată

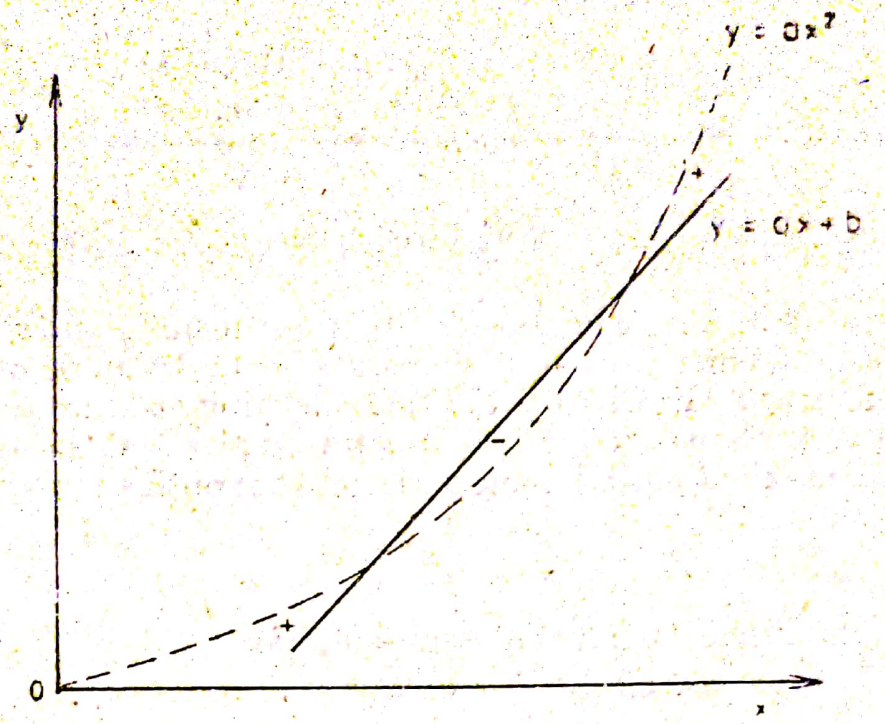


Fig. 4.1. Aproximare unui arc de parabolă printr-o relaţie liniară.

de erori, va fi o prognoză care se abate din ce în ce mai mult de la starea reală a fenomenului respectiv.

Din aceste motive va trebui ca în polinomul de aproximare să luăm încă un termen, termenul $A_2 p_2(x)$.

III. Se are în vedere relaţia

$$p_2(x) = x^2 - \frac{(x^2 p_0)}{(p_0 p_0)} p_0(x) - \frac{(x^2 p_1)}{(p_1 p_1)} p_1(x)$$

unde

$$(x^2 p_1) = 0{,}2210,$$

celelalte mărimi fiind calculate mai sus.

Prin urmare,

$$p_2(x) = x^2 - 20{,}0909\ x + 100{,}9100$$

Avînd în vedere structura tabelei 4.2, scriem

$$(y p_2) = (x^2 p_2) = (p_2 p_2)$$

astfel că

$$A_2 = \frac{(y p_2)}{(p_2 p_2)} = 1.$$

În modul acesta, termenul calculat aici va fi

$$A_2 p_2(x) = x^2 - 20{,}0909\ x + 100{,}9100$$

Scanned with OKEN Scanner

iar polinomul (4.8) devine

$$\begin{array}{r} y = x^2 - 20{,}0909\ x + 100{,}9100 \\ + 20{,}0909\ x - 100{,}9100 \\ \hline y = x^2 \end{array}$$

ceea ce înseamnă că am regăsit formula parabolei de la care am plecat.

Prin urmare, luînd în considerare şi termenul $A_2 p_2(x)$, am ajuns la funcţia dată, chiar dacă arcul de parabolă era destul de mic. Această situaţie ne arată că pentru problemele de prognoză, unde avem de-a face cu extrapolarea, trebuie ca funcţia empirică să fie determinată cu precizia maximă posibilă, numai astfel vom obţine un rezultat demn de încredere.

4.4. Stabilirea formulelor semi-empirice Metoda corecţiilor diferenţiale

La începutul paragrafului 4.1 am arătat că există situaţii în care putem dispune de formule matematice prin care dependenţa lui y de x este exprimată cu ajutorul unor constante. În astfel de împrejurări putem scrie

$$y = f(x,\ a_1,\ a_2,\ \ldots,\ a_m) \tag{4.28}$$

unde $a_i (i = \overline{1,m})$ urmează a fi determinate prin intermediul datelor experimentale x_k, y_k. Abia după aceea relaţia (4.28) se va putea utiliza în problemele de prognoză pentru fenomenul respectiv.

Dacă funcţia (4.28) este de forma (4.4) sau se poate aduce la această formă, prin metoda celor mai mici pătrate se pot determina constantele a_i. Dar dacă (4.28) nu poate fi adusă la forma (4.4), pentru determinarea mărimilor a_i trebuie să căutăm un alt procedeu.

Să presupunem, aşadar, că din analiza graficului

$$y_k = f(x_k)$$

sau din consideraţii teoretice dispunem de nişte valori aproximative $\tilde{a}_i$. În astfel de situaţii, utilizînd metoda corecţiilor diferenţiale, vom determina corecţiile δa_i astfel încît

$$a_i = \tilde{a}_i + \delta a_i \tag{4.29}$$

Cu ajutorul mărimilor aproximative ale constantelor respective, din (4.28), pentru un anumit x_k, avem

$$(y_c)_k = f(x_k,\ \tilde{a}_1,\ \tilde{a}_2,\ \ldots,\ \tilde{a}_m) \tag{4.30}$$

Scanned with OKEN Scanner

Dacă ținem seamă de (4.29) și efectuăm diferența dintre (4.28) și (4.30), putem scrie

$$y_k - (y_c)_k = f(x_k, \tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_m) + \sum_{i=1}^{m} \frac{\partial f}{\partial a_i} \delta a_i - f(x_k, \tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_m)$$

Aici corecțiile δa_i au fost considerate destul de mici astfel încît, în dezvoltarea Taylor, să neglijăm termenii de ordinul doi și mai mare.

Prin urmare,

$$\Delta y_k = \frac{\partial f}{\partial a_1} \delta a_1 + \frac{\partial f}{\partial a_2} \delta a_2 + \ldots + \frac{\partial f}{\partial a_m} \delta a_m \qquad (4.31)$$

unde Δy_k este diferența dintre y_k experimental și cel calculat cu mărimile aproximative $\tilde{a}_i$ introduse în (4.30).

Din datele experimentale, pentru fiecare pereche de valori x_k, y_k se stabilește cîte o ecuație de condiție de forma (4.31) cu necunoscutele δa_i. Rezolvarea unui astfel de sistem, cu mai multe ecuații decît necunoscute, se poate efectua prin metoda celor mai mici pătrate cu luarea în considerare a uneia din schemele amintite în paragraful 4.2.

Observație. Rapoartele $\partial f/\partial a_i$ reprezintă valorile numerice ale derivatelor parțiale, pentru calculul acestora fiind utilizate mărimile aproximative $\tilde{a}_i$. De aceea, rapoartele respective ar putea să fie influențate de faptul că se utilizează mărimi aproximative în locul celor corecte.

Pentru a înlătura, pe cît posibil, influența negativă amintită mai sus, la rezolvarea sistemului (4.31) trebuie să utilizăm o serie de aproximații succesive.

Aplicație practică

Pentru ilustrarea practică a metodei prezentate mai sus, am apelat la datele numerice din primele două coloane ale tabelei 24 (vezi Todoran, 1976, p. 149). Așa cum arătam și în lucrarea amintită, datele respective puteau fi reprezentate printr-o formulă de forma

$$y = a_1 + a_2 \cdot a_3^x \qquad (4.32)$$

Printr-un procedeu destul de anevoios, în lucrarea amintită, am determinat valorile aproximative: $\tilde{a}_1 = 10$, $\tilde{a}_2 = 3$ și $\tilde{a}_3 = 2$. Ne propunem acum ca, prin metoda celor mai mici pătrate și aceea a corecțiilor diferențiale, să ameliorăm aceste valori. În acest scop, din (4.32) putem scrie

$$\frac{\partial f}{\partial a_1} = 1, \qquad \frac{\partial f}{\partial a_2} = a_3^x, \qquad \frac{\partial f}{\partial a_3} = a_2 \cdot x \cdot a_3^{x-1} \qquad (4.33)$$

Pentru stabilirea ecuațiilor de condiție, întocmim tabela 4.3

Tabela 4.3

x	y	$\frac{\partial f}{\partial a_1}$	$\frac{\partial f}{\partial a_2}$	$\frac{\partial f}{\partial a_3}$	$(y_c)_I$	Δy
1	16	1	2	3	16	0
2	22	1	4	12	22	0
3	33	1	8	36	34	−1
4	57	1	16	96	58	−1
5	107	1	32	240	106	+1
6	201	1	64	576	202	−1
7	395	1	128	1344	394	+1
8	777	1	256	3072	778	−1
9	1546	1	512	6912	1546	0

Datele numerice din tabela (4.3) au fost introduse în (4.31) și a rezultat următorul sistem de ecuații de condiție

$$\begin{aligned}
1\,\delta a_1 + 2\,\delta a_2 + 3\,\delta a_3 &= 0\\
1 + 4 + 12 &= 0\\
1 + 8 + 36 &= -1\\
1 + 16 + 96 &= -1\\
1 + 32 + 240 &= +1\\
1 + 64 + 576 &= -1\\
1 + 128 + 1344 &= +1\\
1 + 256 + 3072 &= -1\\
1 + 512 + 6912 &= 0
\end{aligned}$$

De aici a rezultat următorul sistem de ecuații normale

$$\begin{aligned}
9\,\delta a_1 + 1022\,\delta a_2 + 12291\,\delta a_3 &= -2\\
1022\,\delta a_1 + 349524\,\delta a_2 + 4543830\,\delta a_3 &= -184\\
12291\,\delta a_1 + 4543830\,\delta a_2 + 59419305\,\delta a_3 &= -2196
\end{aligned}$$

cu soluțiile:

$$\delta a_1 = -0{,}285, \quad \delta a_2 = 0{,}0035, \quad \delta a_3 = -0{,}00025$$

Înlocuind rezultatele numerice în (4.29) obținem

$$\begin{aligned}
a_1 &= 10 - 0{,}28 = 9{,}72\\
a_2 &= 3 + 0{,}00 = 3{,}00\\
a_3 &= 2 + 0{,}00 = 2{,}00
\end{aligned}$$

Pentru rezolvarea sistemului de ecuații normale am utilizat metoda lui Gauss, de unde a rezultat faptul că la rezultatele obținute nu este cazul să luăm în considerare mai mult de două cifre după virgulă.

Compararea rezultatelor găsite aici cu cele determinate anterior arată că, prin metode diferite, s-au obținut aceleași valori pentru

Scanned with OKEN Scanner

constantele respective. În plus, metoda corecțiilor diferențiale are un domeniu de aplicabilitate mult mai larg, se poate programa la calculator și, la nevoie, aplicînd metoda lui Gauss, ne furnizează informații și asupra preciziei cu care s-au obținut rezultatele respective.

Metoda corecțiilor diferențiale se poate aplica la o clasă largă de funcții și, în final, se bazează și ea pe principiul celor mai mici pătrate. Și aici se face uz de ecuațiile de condiție al căror număr trebuie să fie mai mare decît numărul necunoscutelor.

4.5. Interpolarea și extrapolarea funcțiilor empirice

Să considerăm seria de date experimentale

$$x_1, x_2, \ldots, x_n$$
$$y_1, y_2, \ldots, y_n$$

unde pentru un anumit x_i dispunem de mărimea corespunzătoare y_i.

În cadrul problemelor de prognoză, atunci cînd nu cunoaștem destul de bine legile care stau la baza manifestării unui fenomen dat, se ivesc cazuri în care pentru un anumit x_k avem nevoie de y_k de care nu dispunem din observații sau măsurări. În astfel de situații trebuie să utilizăm metode de interpolare pentru cazul cînd x_k se găsește cel puțin între două argumente x_i și x_{i+1} pentru care dispunem de mărimile y_i și y_{i+1} și metode de extrapolare atunci cînd x_k se află la stînga sau la dreapta argumentelor pentru care dispunem de date empirice. Asupra acestor două categorii de probleme vom insista în prezentul paragraf.

a) *Interpolarea funcțiilor empirice*

Problema interpolării unei funcții empirice este foarte simplă atunci cînd dispunem de formule empirice care au fost obținute prin unul dintre procedeele menționate mai sus. Dar, adeseori, fie că nu putem stabili astfel de formule, fie că problema devine prea complicată pentru efectuarea tuturor calculelor necesare. De aceea, în unele situații este mai comod a dispune de un mijloc de interpolare chiar fără a deduce formula empirică corespunzătoare.

În general, problemele de interpolare sînt expuse în literatura „matematică" privind *analiza numerică* unde sînt bine dezvoltate atît din punct de vedere teoretic cît și practic. Dar, avînd în vedere scopul urmărit aici — utilizarea unor date experimentale pentru probleme de prognoză, interpolarea o vom trata ținînd seamă atît de precizia datelor de care dispunem cît și de precizia rezultatelor pe care le căutăm.

i) Interpolarea grafică și liniară

În cazul cînd dispunem de un număr mare de date experimentale din care rezultă o serie de perechi de valori $x_i y_i$, de obicei se efectuează reprezentarea grafică și se obține cîmpul de distribuție al acestora. Dacă între aceste puncte există o dependență funcțională, adică dacă coeficientul de corelație sau raportul de corelație au valorile apropiate de unitate, se desenează cu mîna liberă curba care trece printre punctele respective. În modul acesta avem apoi posibilitatea ca, pentru orice valoare x_k, să putem citi mărimea corespunzătoare y_k. Acest procedeu este cunoscut sub denumirea de *interpolare grafică*.

Rezultatele obținute prin interpolarea grafică sînt afectate de erorile accidentale ale datelor experimentale disponibile, la care se mai adaugă erorile comise cu ocazia trasării graficului respectiv, precum și acele specifice modului de utilizarea a graficelor în general.

Interpolarea grafică este foarte comodă și ușor de aplicat, dar pentru ca ea să fie cît mai corect posibilă trebuie să se țină seamă de o serie de reguli privind trasarea unei curbe empirice, problemă asupra căreia vom insista mai mult în capitolul următor.

În cazul cînd problemele urmărite cer o interpolare mai riguroasă, vom utiliza expresii matematice stabilite în acest scop. Astfel, dacă punctele definite de datele experimentale sînt destul de dese și prezintă o dependență funcțională evidentă, arcul de curbă dintre două puncte consecutive se poate aproxima printr-un segment de dreaptă. Este vorba de *interpolarea liniară* între punctele empirice. Fie M_i și M_{i+1} două puncte pentru care se cunosc coordonatele (x_i, y_i) și (x_{i+1}, y_{i+1}), iar pentru x_k se caută mărimea y_k, unde $x_i < x_k < x_{i+1}$. În aceste condiții presupunem că avem de-a face cu trei puncte coliniare care definesc sistemul de trei ecuații cu necunoscutele a_0 și a_1,

$$\left.\begin{aligned} y_k &= a_0 + a_1 x_k \\ y_i &= a_0 + a_1 x_i \\ y_{i+1} &= a_0 + a_1 x_{i+1} \end{aligned}\right\} \qquad (4.34)$$

Din ultimele două ecuații din (4.34) putem scrie

$$a_1 = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}$$

astfel încît

$$y_k = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}(x_k - x_i) + y_i \qquad (4.35)$$

Prin urmare, dacă avem două puncte determinate pe cale experimentală, interpolarea mărimii y_k pentru $x_k \in (x_i, x_{i+1})$ se efectuează prin intermediul expresiei (4.35).

Scanned with OKEN Scanner

Dacă dispunem de un număr mare de date (x_i, y_i) astfel că desimea lor, pe un interval, este mare, pentru $x_{i-1} < x_i < x_{i+1} < x_{i+2}$ se poate stabili de asemenea o formulă liniară. În acest scop vom scrie patru ecuații de forma (4.35), fiecare dintre acestea fiind bazată pe cîte un punct din stînga și altul din dreapta lui $M_k(x_k, y_k)$:

$$\left.\begin{aligned}
(y_k)_{i,i+1} &= y_i + \frac{y_{i+1}-y_i}{x_{i+1}-x_i}(x_k - x_i)\\
(y_k)_{i,i+2} &= y_i + \frac{y_{i+2}-y_i}{x_{i+2}-x_i}(x_k - x_i)\\
(y_k)_{i-1,i+1} &= y_{i-1} + \frac{y_{i+1}-y_{i-1}}{x_{i+1}-x_{i-1}}(x_k - x_{i-1})\\
(y_k)_{i-1,i+2} &= y_{i-1} + \frac{y_{i+2}-y_{i-1}}{x_{i+2}-x_{i-1}}(x_k - x_{i-1})
\end{aligned}\right\} \qquad (4.36)$$

Dacă cele patru valori y_k date de relațiile (4.36) sînt la fel de probabile vom calcula media aritmetică a acestora. Prin urmare,

$$y_k = \frac{y_i + y_{i-1}}{2} + \left[\frac{y_{i+1}-y_i}{x_{i+1}-x_i} + \frac{y_{i+2}-y_i}{x_{i+2}-x_i}\right]\frac{x_k - x_i}{4} + \\ + \left[\frac{y_{i+1}-y_{i-1}}{x_{i+1}-x_{i-1}} + \frac{y_{i+2}-y_{i-1}}{x_{i+2}-x_{i-1}}\right]\frac{x_k - x_{i-1}}{4} \qquad (4.37)$$

Dacă avem de-a face cu tabele cu pas constant, h, putem scrie

$$x_{i+1} - x_i = h, \quad x_{i+2} - x_i = 2h, \quad x_{i+1} - x_{i-1} = 2h, \quad x_{i+2} - x_{i-1} = 3h$$

$$x_{i-1} = x_i - h,$$

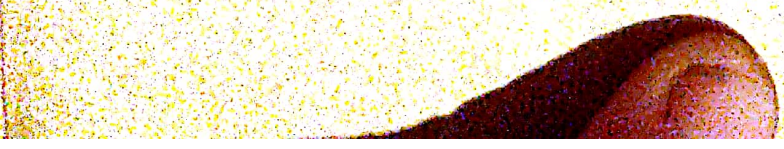
iar formula (4.37) devine

$$y_k = \frac{7(y_i + y_{i-1}) + 3(y_i + y_{i+1}) + 2(y_i + y_{i+2})}{24} + \\ + [9(y_{i+1} - y_i) + 5(y_{i+2} - y_{i-1})]\frac{x_k - x_i}{24\,h} \qquad (4.38)$$

În cazul cînd x_k se găsește la mijlocul intervalului $x_{i+1} - x_i$ avem

$$y_k = \frac{3y_{i-1} + 5y_i + 5y_{i+1} + 3y_{i+2}}{16} \qquad (4.39)$$

După cum se vede, valoarea y_k furnizată de (4.39) este o medie ponderată în care ponderea valorilor externe este 3 iar a celor interne este 5. Cu alte cuvinte, punctele mai apropiate de argumentul pentru care se interpolează aduc o mai mare contribuție la valoarea medie.



Scanned with OKEN Scanner

Aplicație

În tabela 4.4 am reprodus după Hrenov (1951, p. 166) cinci valori ale argumentului x și funcția corespunzătoare sin x. Aici $h = 1'00''$ și ne propunem să interpolăm funcția sin 5°43′00″, mărime care, în conformitate cu Hrenov, are valoarea dată în paranteză. În coloana p_L avem ponderile pentru interpolarea liniară. Din aplicarea formulei (4.39) rezultă

$$y_k = \sin 5°43'00'' = 0{,}09960917.$$

Din compararea lui y_k determinată aici și valoarea din tabelă se vede că diferența este de trei unități la cea de-a opta zecimală. Prin urmare, pentru $5°40'00'' < x_k < 5°46'00''$ cu pasul $h = 2'00''$, formula (4.39) asigură valoarea lui sin x_k cu cel puțin șapte zecimale exacte.

Dacă cele patru puncte considerate mai sus nu se așază pe o linie dreaptă, putem considera că arcul de curbă corespunzătoare se poate aproxima printr-o parabolă de forma

$$y_k = a_0 + a_1 x + a_2 x^2, \quad (x_{i-1} < x_i < x_k < x_{i+1} < x_{i+2}) \tag{4.40}$$

unde a_0, a_1 și a_2 sînt mărimi constante care se pot determina din sistemul

$$\left.\begin{aligned} y_i &= a_0 + a_1 x_i + a_2 x_i^2 \\ y_{i+1} &= a_0 + a_1 x_{i+1} + a_2 x_{i+1}^2 \\ y_{i+2} &= a_0 + a_1 x_{i+2} + a_2 x_{i+2}^2 \end{aligned}\right\} \tag{4.41}$$

sau

$$\left.\begin{aligned} y_i &= a_0 + a_1 x_i + a_2 x_i^2 \\ y_{i-1} &= a_0 + a_1 x_{i-1} + a_2 x_{i-1}^2 \\ y_{i+1} &= a_0 + a_1 x_{i+1} + a_2 x_{i+1}^2 \end{aligned}\right\} \tag{4.42}$$

unde este ușor de văzut că am alcătuit două combinații ale celor patru puncte existente.

Considerînd că prin fiecare grup de cîte trei puncte se definește o parabolă de forma (4.40), rezolvarea sistemelor (4.41) și (4.42) duce la două serii de valori pentru mărimile a_0, a_1 și a_2. Pe acestea le înlocuim în (4.40) și obținem expresiile a două parabole (vezi Todoran 1976, pg. 158—161) și, luînd valoarea medie, avem

$$y_k = y_i + \frac{x_k - x_i}{2}\left\{\frac{y_{i+1} - y_i}{x_{i+1} - x_i}\left[\frac{x_k - x_{i+2}}{x_{i+1} - x_{i+2}} + \frac{x_k - x_{i-1}}{x_{i+1} - x_{i-1}}\right] + \right. \tag{4.43}$$

$$\left. + \left[\frac{y_{i-1} - y_i}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} - \frac{y_{i+2} - y_i}{(x_{i+2} - x_i)(x_{i+1} - x_{i+2})}\right](x_k - x_{i+1})\right\}$$

Scanned with OKEN Scanner

Pentru cazul cînd tabelele au pas constant $(x_{i+1} - x_i = h)$ obţinem

$$y_k = y_i + \frac{x_k - x_i}{4h^2}[(y_{i-1} - y_i - y_{i+1} + y_{i+2})(x_k - x_i) - (y_{i-1} - 5y_i - 5y_{i+1} + y_{i+2})h] \quad (4.44)$$

Dacă trebuie să interpolăm pentru x_k situat chiar la mijlocul intervalului $x_{i+1} - x_i$,

$$y_k = \frac{9(y_{i+1} + y_i) - (y_{i+2} + y_{i-1})}{16} \quad (4.45)$$

Deoarece am folosit aici două arce de parabolă, acest fel de interpolare îl numim interpolare *dublă pătratică*.

Este uşor de văzut că şi formula (4.45) este un fel de medie ponderată, numai că, în comparaţie cu interpolarea liniară (4.39), aici ordonatele y_{i+1} şi y_i au o contribuţie mult mai mare.

Tabela 4.4

x_k	sin x_k	p_L	p_L sin x_k
5°40′00″	0,09874083	3	0,29622249
5 42 00	0,09931975	5	0,49659875
5 43 00	(0,09960920)		
5 44 00	0,09989863	5	0,49949315
5 46 00	0,10047748	3	0,30143244
			1,59374683

Diferenţa dintre (4.45) şi (4.39) este

$$\Delta y_k = \frac{(y_i - y_{i-1}) - (y_{i+2} - y_{i+1})}{4} \quad (4.46)$$

care este nulă numai atunci cînd funcţia empirică, pe intervalul considerat, are o formă liniară, adică

$$y_i - y_{i-1} = y_{i+2} - y_{i+1}$$

Aplicaţie practică pentru interpolarea liniară şi cea dublă pătratică

În tabela 4.5 am reprodus, după Hrenov (1951, p. 263), cinci valori ale argumentului x_k şi ale funcţiei sin x_k. Aici însă liniaritatea

Tabela 4.5

x_k	sin x_k	p_P	p_P sin x_k	p_L	p_L sin x_k
68°	0,9271839	−1	−0,9271839	3	2,7815517
70	0,9396926	9	8,4572334	5	4,6984630
71	(0,9455186)				
72	0,9510565	9	8,5595085	5	4,7552825
74	0,9612617	−1	−0,9612617	3	2,8837851
			15,1282963		15,1190823

funcției este greu de admis. Semnificația coloanelor din tabela 4.5 este specificată, ponderile p_P și p_L fiind specifice pentru interpolarea pătratică și respectiv cea liniară.

Pentru interpolarea dublă pătratică efectuăm înlocuirile în (4.45) și avem

$$(y_k)_P = \sin 71^\circ = \frac{15,1282963}{16} = 0,9455185$$

pe cînd pentru interpolarea liniară rezultă

$$(y_k)_L = \sin 71^\circ = \frac{15,1190823}{16} = 0,9449443$$

Comparația dintre aceste mărimi și sinusul respectiv din tabele arată că interpolarea dublă pătratică dă un rezultat foarte bun, pe cînd interpolarea liniară, pe acest interval, nu este potrivită.

b) *Extrapolarea funcțiilor empirice*

După cum am mai amintit, în problemele de prognoză avem de-a face cu cercetarea unei funcții empirice în domeniul definit de abscisele situate la stînga sau la dreapta intervalului în care dispunem de date experimentale. În astfel de situații trebuie să determinăm anumite ordonate pentru abscisele la care nu avem date experimentale și care sînt situate în afara intervalului în care dispunem de astfel de ordonate. De aceea, în problemele de prognoză avem de-a face cu așa numita problemă de extrapolare.

Problema extrapolării unei funcții empirice se poate rezolva prin utilizarea unor formule empirice sau semi-empirice stabilite prin unul din procedeele expuse anterior. Dar și în astfel de cazuri trebuie să fim siguri că funcția respectivă se va comporta și în viitor la fel ca și pînă acum.

Utilizarea formulelor empirice pentru probleme de extrapolare va asigura o prognoză bună numai atunci cînd stabilirea formulei empirice se bazează atît pe date observaționale numeroase, cît și pe cunoașterea legilor care stau la baza manifestării fenomenului examinat. Exemple de acest fel sînt multe: prevederea modului de desfășurare a eclipselor de Lună și de Soare, prevederea poziției planetelor pe bolta cerească pentru un moment dat, determinarea poziției reciproce a componentelor unei stele duble etc.

În cazul cînd nu cunoaștem legile care stau la baza desfășurării unui fenomen dat, cum ar fi, de exemplu, legile mecanicii cerești pentru fenomenele amintite mai sus, formulele empirice se deduc numai pe baza datelor experimentale. Și în acest caz putem dispune de o extrapolare utilă, însă, în astfel de situații, trebuie să urmărim cu multă grijă comportarea pe viitor a acestor funcții. Astfel, de exemplu, dacă dispunem de o serie de observații distribuite pe un

Scanned with OKEN Scanner

interval mic de timp, s-ar putea ca acestea să definească o dependență liniară, să reprezinte un arc de parabolă sau să facă parte dintr-o curbă definită printr-o funcție periodică. De aceea, supravegherea atentă a comportării viitoare a funcției respective este foarte necesară. Este absolut necesar să urmărim modul în care prevederea este în concordanță cu observația și, la momentul cînd se constată că datele observaționale se abat mult de la valorile prevăzute, să se refacă calculele și să fie restabilită o nouă formulă empirică care să verifice toate datele experimentale, să fie în acord cu noile observații.

Menționăm aici că niște metode asemănătoare au stat la baza descoperirilor teoretice ale planetelor Neptun și Pluto. Aici, deși se cunoșteau legile de mișcare ale corpurilor cerești, prevederile privind pozițiile planetei Uranus erau afectate de perturbații. Se puneau în evidență diferențe sistematice între pozițiile prevăzute și cele observate.

Pornind de la constatarea diferențelor dintre prevederi și realitate, s-a presupus că există o cauză a acestor deviații, că ar exista o planetă necunoscută care perturbă mișcarea lui Uranus. Tratînd, apoi în mod corect această legătură de cauzalitate, s-a ajuns la o prevedere teoretică, a poziției planetelor necunoscute poziții pe care astronomii observatori le-au confirmat.

În cazul cînd nu dispunem de date suficiente pentru a stabili formule empirice, putem utiliza graficele curbelor respective. În astfel de situații extrapolarea se va baza pe trasarea cu mîna liberă a curbei în afara punctelor existente pe graficul respectiv. Acest procedeu nu este destul de precis, dar pentru o prognoză pe termen scurt, el poate aduce rezultate deosebit de importante. Pe această cale se pot prevedea momentele cînd va avea loc intensitatea maximă sau minimă a manifestării unor fenomene.

Scanned with OKEN Scanner

CAPITOLUL V

UTILIZAREA GRAFICELOR PENTRU PROBLEME DE PROGNOZĂ

Din activitatea practică se știe că, atunci cînd se elaborează studiul unui fenomen natural sau al unei activități sociale oarecare, analiza anumitor grafice poate aduce o contribuție substanțială atît în privința examinării modului de desfășurare a acestora în timp, cît și în ceea ce privește unele probleme de extrapolare care ar putea sta la baza prevederii desfășurării în viitor a fenomenului studiat — probleme de prognoză. Prin urmare, anumite reprezentări grafice pot fi de mare utilitate în problemele de prognoză, dar modul în care se construiesc sau se analizează graficele respective va avea o mare influență asupra probabilității ca fenomenul analizat să se desfășoare în conformitate cu prevederile noastre.

După cum se știe, în activitatea experimentală avem de a face cu date numerice care, pe un grafic, definesc puncte discrete. De aceea, modul în care se va trasa o curbă printre punctele astfel obținute depinde, în mare măsură, de abilitatea cercetătorului, dar, în același timp, există și anumite procedee pe care le putem utiliza în vederea înlăturării, pe cît posibil, a influenței erorilor accidentale. Astfel, pentru ca o curbă empirică să fie corect trasată, ea trebuie să treacă nu direct prin punctele observate ci printre ele. Asupra unor astfel de procedee vom insista în capitolul prezent.

5.1. Importanța reprezentării grafice în problemele de corelație și cauzalitate

În capitolul precedent am văzut că problemele de prognoză sînt mult simplificate dacă dispunem de o serie mare de date experimentale, pe baza cărora se deduc formule empirice sau se stabilesc modele matematice. În astfel de cazuri, relația matematică stabilită ne arată natura dependenței dintre anumite variabile precum și modul de manifestare a fenomenului studiat. De aici se pot trage concluzii asupra probabilității desfășurării viitoare a fenomenelor sau proceselor studiate.

În cazul cînd, utilizînd formula empirică stabilită, punem în evidență legile principale care stau la baza anumitor fenomene, putem întocmi modele fizice pe baza cărora prognoza matematică devine

Scanned with OKEN Scanner

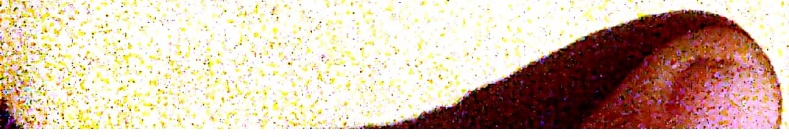
mult mai demnă de încredere. Dar, spre regretul nostru, astfel de situaţii sînt destul de rare şi, în marea majoritate a cazurilor, noi dispunem de datele observaţionale din care trebuie să stabilim o prognoză fie ea chiar şi cu un mic grad de probabilitate. Există situaţii în care reprezentarea grafică a datelor observaţionale şi analiza acestor grafice sînt singurele mijloace de care dispunem atît pentru stabilirea unor corelaţii existente cît şi pentru suspectarea sau detectarea anumitor cauzalităţi. În plus, în anumite cazuri, examinarea grafică poate fi utilizată pentru detectarea imediată a prezenţei unor cauzalităţi pentru studiul cărora va trebui să stabilim noi planuri de cercetare, obţinerea de noi date observaţionale şi reluarea problemei pe o bază mult mai fundamentată. Astfel, de exemplu, dacă la o stea se observă un început de creştere considerabilă a strălucirii, cum este cazul novelor sau al supernovelor, observatorul respectiv îşi îndreaptă mai mult atenţia asupra acestui obiect şi, eventual, constatarea sa o comunică şi la alţi observatori. În modul acesta se efectuează observaţii mult mai numeroase; se adună un vast material experimental şi abia după ce fenomenul respectiv a încetat să mai prezinte interes, se reiau datele observaţionale, se reprezintă grafic şi, după forma şi caracteristicile curbei de variaţie a strălucirii, se stabilesc valorile pentru o serie de parametri fizici care au stat la baza declanşării fenomenului respectiv (nova sau supernova în cazul amintit aici).

Dacă un astfel de fenomen a fost observat pe un interval scurt de timp, reprezentarea grafică a datelor experimentale ne poate indica doar clasa de fenomene printre care se poate încadra şi fenomenul observat de noi. În cazul amintit mai sus, după rata creşterii strălucirii se poate decide dacă este vorba de o novă sau o supernovă şi, apoi, eventual se poate spune despre ce fel de novă sau ce fel de supernovă este vorba. Prin urmare, o reprezentare grafică a datelor experimentale, în faza de început a desfăşurării unui fenomen dat, poate servi la o şi mai bună planificare a muncii de cercetare.

În cazul cînd, din anumite motive, nu putem efectua serii sistematice de observaţii, prin compararea dintre graficul observat şi graficele corespunzătoare ale altor fenomene de acelaşi fel, se pot trage concluzii asupra categoriei de fenomene din care face parte şi cel observat de noi. Astfel, de exemplu, în figura 5.1a sînt reprezentate cîteva înregistrări efectuate sporadic la o stea variabilă, iar pe figura 5.1b se găsesc curbele de variaţie a strălucirii la cele trei tipuri de stele variabile din clasa al cărei reprezentant este steaua RR Lyrae.

Dacă copiem pe hîrtie de calc datele din figura 5.1a şi le suprapunem peste curbele din figura 5.1b, se vede imediat că fenomenul observat este legat de variaţia strălucirii unei stele de tipul RR Lyrae, subclasa c.

Dacă este vorba de un cutremur de Pămînt, erupţia unui vul-



Scanned with OKEN Scanner

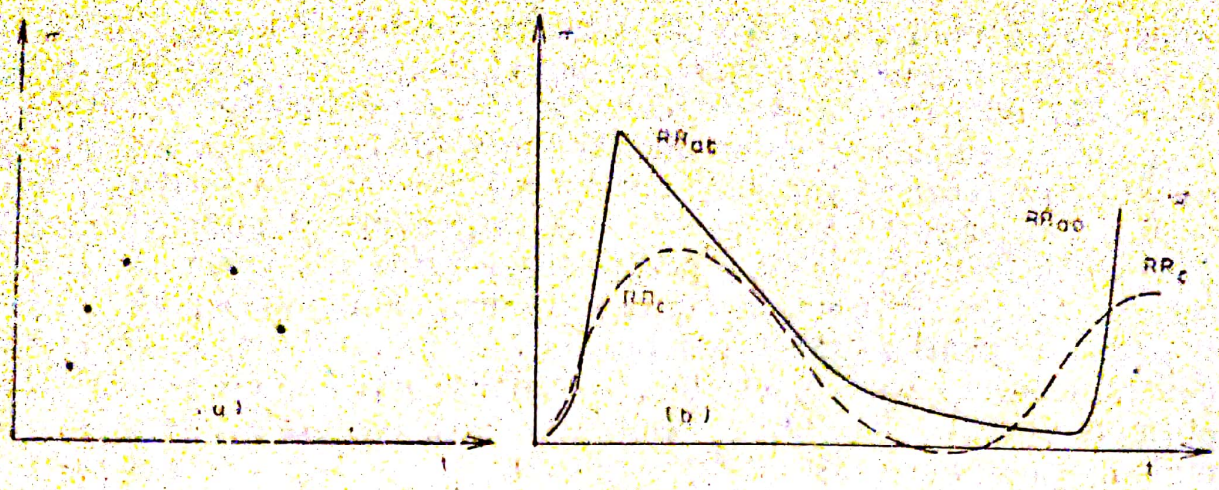


Fig. 5.1. Comparaţia dintre variaţia observată a strălucirii unei stele variabile şi curbele corespunzătoare ale stelelor de tipul RR Lyrae.

can, ritmul de creştere al cotelor apelor unui fluviu etc, analiza grafică a primelor date observaţionale poate să scoată în evidenţă atît tăria fenomenului pe care-l urmărim cît şi modul în care urmează să se desfăşoare mai departe. În felul acesta se pot lua măsurile de reducere la minim a acţiunilor distrugătoare (erupţii vulcanice, cutremure de Pămînt, epidemii etc) sau, eventual, de a lua măsurile necesare pentru ca fenomenele respective să fie utilizate pentru binele nostru (creşterea debitului unor rîuri, declanşarea vînturilor etc).

În cazul unor fenomene periodice, primele date experimentale ar putea să indice un grafic din care să rezulte atît amplitudinea cît şi perioada fenomenului respectiv. Prin urmare, o primă analiză a datelor experimentale poate da informaţii preţioase cu privire la etapele care urmează în desfăşurarea evenimentelor şi, apoi, să se întreprindă studii mai detaliate pe baza analizei numerice a datelor respective. În orice caz, reprezentarea grafică poate scoate în evidenţă corelaţia existentă — corelaţia vădită — care este determinată de o cauză bine definită sau cauze multiple care se pot suprapune şi să indice o manifestare foarte complicată a fenomenului respectiv.

Există situaţii în care reprezentarea grafică este singurul mijloc eficient şi rapid din care se pot trage concluzii cu privire la corelaţii, cauzalităţi şi probleme de prognoză. Tocmai din acest motiv, capitolul prezent este consacrat expunerii şi examinării unor procedee prin intermediul cărora să obţinem maximul de informare într-un timp cît mai scurt. În acest scop, atunci cînd este foarte dificil sau imposibil să obţinem o formulă empirică prin care să reprezentăm datele observaţionale, pentru a avea totuşi o privire generală asupra dependenţei funcţionale $y = f(x)$, vom executa reprezentarea grafică a punctelor observate $y_i = f(x_i)$.

Pentru a avea o idee asupra dependenţei funcţionale $y = f(x)$, vom trasa o curbă printre punctele observate, dar această curbă să

Scanned with OKEN Scanner

fie dusă în așa fel încît de o parte și de alta a ei să rămînă aproximativ același număr de puncte. Trasarea unei astfel de curbe, dacă ea este corectă, poate fi utilizată în locul unei formule empirice, pentru orice abscisă x_i să putem citi ordonata y_i sau să tragem anumite concluzii asupra fenomenului cercetat cum ar fi, de exemplu, evoluția lui în viitor — probleme de prognoză.

Deoarece observațiile sînt afectate de erori accidentale care nu se cunosc și nu pot fi eliminate, punctele de pe grafic vor avea o anumită dispersie, vor defini o bandă de o anumită lățime iar curba pe care urmează s-o trasăm va trebui să treacă prin mijlocul acestei benzi. În acest scop vom distinge două categorii de procedee: în cazul cînd dispunem de un mare număr de date experimentele — observații foarte dense, este convenabil să formăm grupe de observații — *puncte normale*, iar curba trasată prin astfel de puncte se va numi *curbă medie*; în al doilea caz, cînd avem observații mai puține, se poate aplica un procedeu de ajustare a datelor observaționale, un procedeu de „corectare" și reducere a dispersiei unei serii de puncte, un procedeu de „netezire" a curbelor empirice. Asupra unor astfel de probleme vom insista în cele ce urmează.

5.2. Formarea punctelor normale. Trasarea curbelor medii

În cazul cînd dispunem de un număr foarte mare de date experimentale care definesc punctele empirice $M_i(x_i, y_i)$, adesea este convenabil a înlocui punctele individuale prin puncte medii; o curbă astfel construită se numește curbă medie.

Dacă pe un anumit domeniu al unei curbe empirice punctele individuale (empirice) indică prezența unei dependențe liniare, coordonatele fiecărui punct trebuie să verifice ecuația unei drepte, inclusiv punctul mediu $M_k(\bar{x}_k, \bar{y}_k)$, adică

$$\bar{y}_k = a_0 + a_1 \bar{x}_k \tag{5.1}$$

unde avem notațiile obișnuite

$$\bar{x}_k = \left(\frac{\sum_{i=1}^{n} x_i}{n} \right)_k, \qquad \bar{y}_k = \left(\frac{\sum_{i=1}^{n} y_i}{n} \right)_k \tag{5.2}$$

Aici k reprezintă numărul de ordine al intervalului din interiorul căruia punctele individuale au fost grupate într-un punct normal folosind în acest scop relațiile (5.2).

Uneori dispunem de un număr foarte mare de observații, iar intervalele în care am considerat dependențele liniare sînt destul de mici. Din cauza dispersiei este greu de hotărît dacă unul sau mai multe puncte trebuie să aparțină uneia sau alteia dintre porțiunile

Scanned with OKEN Scanner

liniare. În astfel de situații, pentru gruparea punctelor individuale în puncte medii, se poate adopta regula de a grupa același număr de puncte în fiecare punct mediu, evident considerîndu-le în ordinea strict crescătoare a absciselor.

Punctele medii, alcătuite în modul prezentat mai sus, pot avea ponderi diferite. Acestea depind de numărul n al punctelor individuale componente precum și de erorile medii pătratice ale mediilor respective $\bar{x}$ și $\bar{y}$. Prin urmare, pentru fiecare punct mediu ponderea va fi definită prin relația

$$p = \frac{n}{w^2}. \tag{5.3}$$

Aici am notat

$$w^2 = (em)_{\bar{x}}^2 + (em)_{\bar{y}}^2, \tag{5.4}$$

erorile medii pătratice ale valorilor medii fiind definite prin

$$(em)_{\bar{x}}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n(n-1)} \tag{5.5}$$

$$(em)_{\bar{y}}^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n(n-1)} \tag{5.6}$$

Aplicație practică

În primele două coloane ale tabelei 5.1 sînt reproduse o serie de date experimentale pe care ne propunem să le grupăm în trei puncte normale. În acest scop am completat celelalte coloane, avînd în vedere cele trei serii de date. Am obținut astfel următoarele rezultate:

$M_1(\bar{x}_1, \bar{y}_1)$ cu $\bar{x}_1 = 38{,}25 \pm 0{,}07$, $\bar{y}_1 = 20{,}00 \pm 0{,}97$, $p_1 = 6$,

$M_2(\bar{x}_2, \bar{y}_2)$ cu $\bar{x}_2 = 38{,}85 \pm 0{,}07$, $\bar{y}_2 = 23{,}83 \pm 1{,}30$, $p_2 = 4$,

$M_3(\bar{x}_3, \bar{y}_3)$ cu $\bar{x}_3 = 39{,}45 \pm 0{,}07$, $\bar{y}_3 = 28{,}17 \pm 1{,}27$, $p_3 = 4$,

Aici, relația (5.3) a fost utilizată sub forma

$$p_k = \frac{n^2(n-1)}{\sum_{i=1}^{n} (x_i - \bar{x}_k)^2 + \sum_{i=1}^{n} (y_i - \bar{y}_k)^2} \tag{5.7}$$

Scanned with OKEN Scanner

Tabela 5.1

x	y		$x - \bar{x}$	$(x - \bar{x})^2$	$y - \bar{y}$	$(y - \bar{y})^2$
38,0	18	$\bar{x}_1 = 38,25$	−0,25	0,0625	−2,00	4,0000
38,1	21		−0,15	,0225	1,00	1,0000
38,2	23	$\bar{y}_1 = 20,00$	−0,05	,0025	3,00	9,0000
38,3	17		0,05	,0025	−3,00	9,0000
38,4	19		0,15	,0225	−1,00	1,0000
38,5	22		0,25	,0625	2,00	4,0000
38,6	26	$\bar{x}_2 = 38,85$	−0,25	,0625	2,17	4,7089
38,7	19		−0,15	,0225	−4,83	23,3289
38,8	25	$\bar{y}_2 = 23,83$	−0,05	,0025	1,17	1,3689
38,9	23		0,05	,0025	−0,83	0,6889
39,0	28		0,15	,0225	4,17	17,3889
39,1	22		0,25	,0625	−1,83	3,3489
39,2	25	$\bar{x}_3 = 39,45$	−0,25	,0625	−3,17	10,0489
39,3	27		−0,15	,0225	−1,17	1,3689
39,4	25	$\bar{y}_3 = 28,17$	−0,05	,0025	−3,17	10,0489
39,5	30		0,05	,0025	1,83	3,3489
39,6	33		0,15	,0225	4,83	23,3289
39,7	29		0,25	,0625	0,83	0,6889

În cazul cînd pe un anumit interval punctele individuale (empirice) prezintă o dependenţă parabolică, pentru fiecare punct putem scrie

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 \tag{5.8}$$

Dacă punem condiţia ca şi punctul mediu să se afle pe aceeaşi parabolă definită de (5.8), avem

$$\bar{y}_k = a_0 + a_1 \bar{x}_k + a_2 (\bar{x})_k^2 \tag{5.9}$$

care este diferită de

$$\bar{y}_k = a_0 + a_1 \bar{x}_k + a_2 (\overline{x^2})_k \tag{5.10}$$

pe care trebuia s-o obţinem dacă am fi procedat ca şi în cazul dependenţei liniare.

Prin urmare, formarea punctelor normale în cazul dependenţelor parabolice este mult mai complicată. Pentru determinarea constantelor a_0, a_1 şi a_2 trebuie rezolvat sistemul de ecuaţii de condiţie (5.8) prin metoda celor mai mici pătrate sau aplicînd altă metodă potrivită. Aşadar, în astfel de situaţii apar dificultăţi deosebite. De aceea, este recomandabil a grupa puncte individuale în puncte medii numai pe acele porţiuni din curbele empirice, unde, în cadrul dispersiei punctelor individuale, arcul de curbă empirică se poate aproxima printr-un segment rectiliniu.

Scanned with OKEN Scanner

5.3. Ajustarea datelor experimentale
Netezirea curbelor empirice

În cazul cînd, din anumite motive, nu ne putem servi de curbe medii, este bine să reducem, pe cît posibil, rolul erorilor accidentale, să micșorăm dispersia, numărul de puncte fiind păstrat nemodificat. În acest sens este vorba de un fel de netezire a curbelor empirice, netezire care se va face în funcție de curbura corespunzătoare.

a) *Trasarea unei curbe empirice cu curbură mică*

Dacă datele experimentale determină puncte care definesc o curbă de curbură mică, adică o curbă care pe porțiuni destul de mici se poate aproxima printr-o linie dreaptă, atunci fiecare trei puncte consecutive se pot grupa împreună pentru a defini un punct mediu. În acest scop vom prezenta aici două procedee : i) procedeul obișnuit de netezire care se bazează pe principiul celor mai mici pătrate, și ii) un procedeu de netezire care se bazează pe faptul că prin fiecare pereche de două puncte se duce o dreaptă și, în felul acesta, se ia valoarea medie a rezultatelor obținute parțial.

i) Procedeul de netezire a unei curbe empirice în care se utilizează metoda celor mai mici pătrate a mai fost discutat de noi (vezi Todoran, 1976, p. 174). Revenim asupra acestuia și-l reexaminăm în vederea unei comparații cu cel de al doilea procedeu pe care-l vom prezenta aici pentru prima dată și pe care-l considerăm mai potrivit din punct de vedere practic.

În cazul cînd pe un anumit interval datele experimentale indică prezența unei variații liniare sau a unei curbe de curbură mică, pe un subinterval destul de mic putem admite că fiecare trei puncte consecutive sînt situate pe o dreaptă. În acest caz, pentru $x \in (x_{i-1}, x_{i+1})$, putem scrie

$$\tilde{y} = a_i x + b_i \tag{5.11}$$

unde a_i și b_i sînt două mărimi constante, iar $\tilde{y}$ reprezintă ordonata socotită pînă la dreapta presupusă că trece prin cele trei puncte observate. Prin urmare, ordonata $\tilde{y}$ reprezintă o ordonată medie care este mai puțin influențată de erorile accidentale decît este ordonata observată corespunzătoare.

În condițiile amintite mai sus, diferențele dintre ordonata empirică (observată), y, și ordonata medie, $\tilde{y}$, trebuie să fie de așa natură încît suma pătratelor diferențelor acestora să fie minimă, adică

$$f(a_i, b_i) = \sum_{j=1}^{3} (a_i x_{ij} + b_i - y_{ij})^2 = \text{minim} \tag{5.12}$$

Scanned with OKEN Scanner

unde i este numărul de ordine al unei secvențe de trei puncte consecutive, iar j reprezintă numărul de ordine al observațiilor din secvența i.

Ținînd seamă de condițiile de minim ale unei funcții date, adică în cazul prezent

$$\frac{\partial f}{\partial a_i} = 0, \qquad \frac{\partial f}{\partial b_i} = 0 \tag{5.13}$$

putem determina mărimile a_i și b_i pe care le înlocuim în (5.11) și, pentru un anumit x, rezultă ordonata medie. În acest scop, sistemul (5.13) devine

$$\sum_{j=1}^{3} (a_i x_{ij} + b_i - y_{ij}) = 0 \tag{5.14}$$

$$\sum_{j=1}^{3} (a_i x_{ij} + b_i - y_{ij}) x_{ij} = 0 \tag{5.15}$$

sau, efectuînd însumările respective, putem scrie mai departe

$$\left.\begin{aligned} 3b_i + a_i \sum_{j=1}^{3} x_{ij} &= \sum_{j=1}^{3} y_{ij} \\ b_i \sum_{j=1}^{3} x_{ij} + a_i \sum_{j=1}^{3} x_{ij}^2 &= \sum_{j=1}^{3} x_{ij} y_{ij} \end{aligned}\right\} \tag{5.16}$$

Sistemul (5.16) conține două necunoscute a_i și b_i, pentru determinarea cărora avem

$$a_i = \frac{\Sigma x_{ij} \Sigma y_{ij} - 3\Sigma y_{ij} x_{ij}}{(\Sigma x_{ij})^2 - 3\Sigma x_{ij}^2} = \frac{Y_{ij} - X_{ij}}{Y'_{ij} - X'_{ij}} \tag{5.17}$$

$$b_i = \frac{1}{3} [\Sigma y_{ij} - a_i \Sigma x_{ij}] \tag{5.18}$$

Avînd în vedere faptul că x_{ij} și y_{ij} sînt mărimi cunoscute, obținute pe cale experimentală, relațiile (5.17) și (5.18) determină valorile constantelor a_i și b_i. Apoi, efectuînd înlocuirile în (5.11) obținem pentru ordonata celui de al doilea punct individual dat

$$\bar{y}_{i+1} = a_i x_{i+1} + b_i, \qquad (i = 1, 2, 3) \tag{5.19}$$

adică obținem ordonata medie $\bar{y}_{i+1}$ care, împreună cu x_{i+1} determină punctul prin care trebuie să treacă curba respectivă.

Scanned with OKEN Scanner

Dacă punctele observate au abscisele echidistante (pas constant) adică dacă $h = x_{i+1} - x_i = \text{constant}$, expresiile (5.17) și (5.18) devin

$$a_i = \frac{y_{i3} - y_{i1}}{2h} \tag{5.20}$$

$$b_i = \frac{1}{3}\left[y_{i1} + y_{i2} + y_{i3} - \frac{y_{i3} - y_{i1}}{2h} \cdot 3x_{i+1}\right] \tag{5.21}$$

Înlocuind pe (5.20) şi (5.21) în (5.19), avem

$$\tilde{y}_{i+1} = \frac{1}{3}(y_{i1} + y_{i2} + y_{i3}) \tag{5.22}$$

sau

$$\tilde{y}_i = \frac{1}{3}(y_{i-1} + y_i + y_{i+1}) \tag{5.23}$$

În cadrul expunerii prezentate mai sus am considerat primele trei puncte observate ($i = 2$) cu ajutorul cărora am determinat ordonata corectată a punctului al doilea. Apoi, lăsăm primul punct observat şi luăm in considerare următoarele trei puncte cu ajutorul cărora vom corecta ordonata celui de al treilea punct observat (al doilea din secvenţa a doua) etc. Procedînd în modul acesta pentru toate combinaţiile de cîte trei puncte observate succesive, putem corecta toate ordonatele cu excepţia primului şi ultimului punct care rămîn necorectate.

Observaţii. a) Dacă funcţia empirică este periodică, adică dacă valorile acesteia se „repetă" după un interval de timp egal cu perioada, P, atunci procedeul descris mai sus se poate aplica pentru toate punctele observate. În acest caz, dacă ne referim la primul punct al primei secvenţe de trei puncte, atunci pentru corectarea ordonatei y_{11} ne folosim de ordonata y_{i0} pe care nu o avem din observaţii, dar în locul ei se va lua ordonata de la faza egală cu $x_{i0} \pm P$. În mod analog se procedează şi pentru ultimul punct observat.

b) Procedeul descris mai sus are cel puţin două neajunsuri:

— s-a aplicat metoda celor mai mici pătrate la un sistem de trei ecuaţii cu două necunoscute, adică numărul ecuaţiilor era numai cu o unitate mai mare decît numărul necunoscutelor,

— pentru rezolvarea sistemului de ecuaţii (5.16) toate ordonatele au intrat cu aceeaşi pondere.

ii) Pentru a expune cel de al doilea procedeu amintit mai sus, vom considera iarăşi grupe de cîte trei puncte, dar acum prin fiecare pereche de cîte două puncte vom duce cîte o dreaptă (vezi Fig. 5.2). În acest caz pentru primele două puncte vom putea scrie ecuaţia generală a unei drepte

$$y = A \cdot x + B \tag{5.24}$$

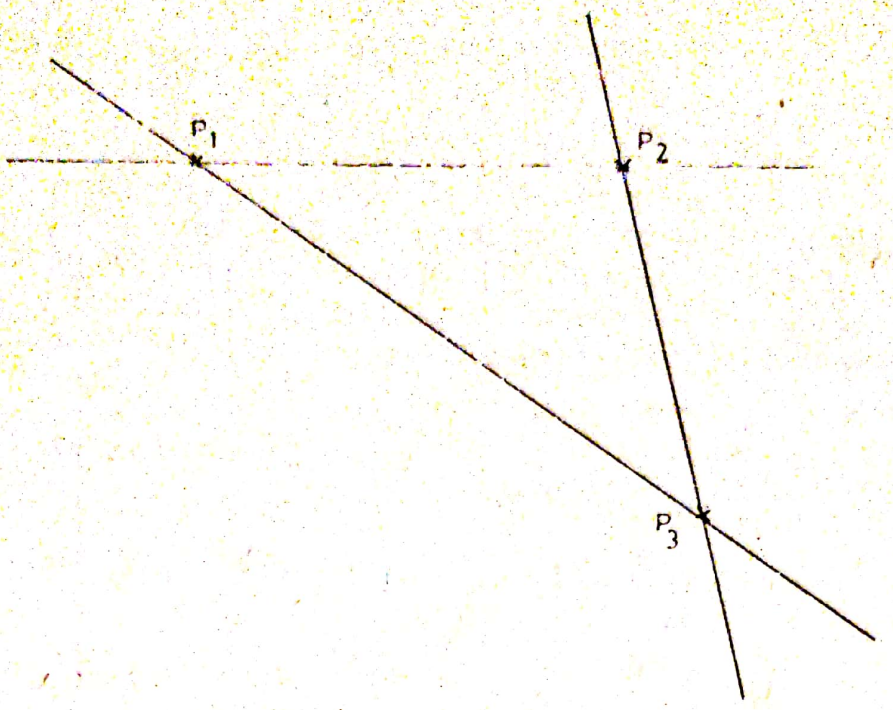


Fig. 5.2. Poziția relativă a celor trei drepte care se pot duce prin trei puncte observate.

care trece prin punctele $P_1(x_1, y)_1$, $P_2(x_2, y_2)$. Prin urmare, putem scrie următorul sistem de ecuații

$$\begin{aligned} y &= A \cdot x + B \\ y_1 &= A \cdot x_1 + B \\ y_2 &= A \cdot x_2 + B \end{aligned} \tag{5.25}$$

de unde avem imediat

$$y_1 - y_2 = A(x_1 - x_2), \qquad A = \frac{y_1 - y_2}{x_1 - x_2} \tag{5.26}$$

În mod analog mai putem scrie

$$\begin{aligned} y - y_1 &= A(x - x_1) \\ y - y_2 &= A(x - x_2) \end{aligned}$$

unde înlocuim expresia coeficientului A și avem

$$y - y_1 = \frac{y_1 - y_2}{x_1 - x_2}(x - x_1); \qquad y - y_2 = \frac{y_1 - y_2}{x_1 - x_2}(x - x_2), \tag{5.27}$$

pe care le adunăm și obținem

$$y = \frac{y_1 + y_2}{2} + \frac{y_1 - y_2}{x_1 - x_2}\left(x - \frac{x_1 + x_2}{2}\right)$$

Scanned with OKEN Scanner

sau

$$y = \frac{1}{x_1 - x_2}[y_1(x - x_2) - y_2(x - x_1)] \tag{5.28}$$

care este forma generală a ecuației dreptei care trece prin punctele $P_1(x_1, y_1)$ și $P_2(x_2, y_2)$, adică cuația dreptei (a) de pe figura 5 .2. În mod analog se procedează și pentru celelalte perechi de puncte, astfel că se obține următorul sistem de ecuații ale celor trei drepte considerate

$$\left.\begin{aligned} &\text{(a)} \quad y = \frac{1}{x_1 - x_2}[y_1(x - x_2) - y_2(x - x_1)] \\ &\text{(b)} \quad y = \frac{1}{x_1 - x_3}[y_1(x - x_3) - y_3(x - x_1)] \\ &\text{(c)} \quad y = \frac{1}{x_2 - x_3}[y_2(x - x_3) - y_3(x - x_2)] \end{aligned}\right\} \tag{5.29}$$

Efectuînd adunarea celor trei relații din (5.29), rezultă

$$3y = y_1\left(\frac{x - x_2}{x_1 - x_2} + \frac{x - x_3}{x_1 - x_3}\right) + y_2\left(\frac{x - x_3}{x_2 - x_3} - \frac{x - x_1}{x_1 - x_2}\right) - y_3\left(\frac{x - x_2}{x_2 - x_3} + \frac{x - x_1}{x_1 - x_3}\right). \tag{5.30}$$

Dacă dorim să „netezim" ordonata y_2, atunci scriem formula (5.30) pentru $x = x_2$, adică

$$\tilde{y}_2 = \frac{2}{3}y_2 + \frac{1}{3(x_3 - x_1)}[y_1(x_3 - x_2) + y_3(x_2 - x_1)] \tag{5.31}$$

care, pentru cazul general, devine

$$\tilde{y}_i = \frac{2}{3}y_i + \frac{1}{3(x_{i+1} - x_{i-1})}[y_{i-1}(x_{i+1} - x_i) + y_{i+1}(x_i - x_{i-1})]. \tag{5.32}$$

Formula (5.32) are avantajul că se poate utiliza într-un caz general, cînd pasul absciselor este variabil. Pentru cazul particular cînd pasul absciselor este constant, adică atunci cînd

$$x_{i+1} - x_i = h = \text{constant},$$

formulele (5.31) și (5.32) devin

$$\tilde{y}_2 = \frac{y_1 + 4y_2 + y_3}{6}, \qquad \tilde{y}_i = \frac{y_{i-1} + 4y_i + y_{i+1}}{6} \tag{5.33}$$

Scanned with OKEN Scanner

În vederea unei comparații dintre formulele (5.23) și (5.33) le scriem una lîngă alta

$$(\tilde{y}_i)_A = \frac{y_{i-1} + y_i + y_{i+1}}{3} \qquad (5.23)$$

$$(\tilde{y}_i)_B \frac{y_{i-1} + 4y_i + y_{i+1}}{6} \qquad (5.33)$$

De aici se vede că în formula (5.23) avem media aritmetică a trei mărimi considerate toate trei de aceeași pondere, pe cînd în formula (5.33) avem o medie ponderată unde punctul care urmează a fi „netezit" are o pondere de patru ori mai mare decît punctele adiacente.

Diferența dintre (5.23) și (5.33) se poate scrie sub forma

$$(\tilde{y}_i)_A - (\tilde{y})_B = \Delta y_i = \frac{(y_{i+1} - y_i) - (y_i - y_{i-1})}{6} \qquad (5.34)$$

unde A și B reprezintă formulele (5.23) și respectiv (5.33). Prin urmare, diferența dintre cele două formule devine nulă numai în cazul în care cele trei puncte sînt strict coliniare.

Aplicație numerică

În tabela 5.2 am reprodus primele coloane din tabela 5.1, unde x și y joacă rolul unor date experimentale. Curba empirică corespun-

Tabela 5.2

x	y	y_{AI}	y_{BI}	y_{AII}	y_{BII}
38,0	18	18,0	18,0	18,0	18,0
38,1	21	20,7	20,8	19,7	20,5
38,2	23	20,3	21,7	20,2	21,0
38,3	17	19,7	18,3	19,8	19,0
38,4	19	19,3	19,2	20,4	19,6
38,5	22	22,3	22,2	21,3	22,0
38,6	26	22,3	24,2	22,6	23,4
38,7	19	23,3	21,2	22,6	22,1
38,8	25	22,3	23,7	23,6	23,4
38,9	23	25,3	24,2	24,0	24,5
39,0	28	24,3	26,2	24,9	25,4
39,1	22	25,0	23,5	24,7	24,2
39,2	25	24,7	24,8	25,1	24,8
39,3	27	25,7	26,3	25,9	26,0
39,4	25	27,3	26,2	27,4	26,8
39,5	30	29,3	29,7	29,1	29,5
39,6	33	30,7	31,8	29,7	31,0
39,7	29	29,0	29,0	29,0	29,0

Scanned with OKEN Scanner

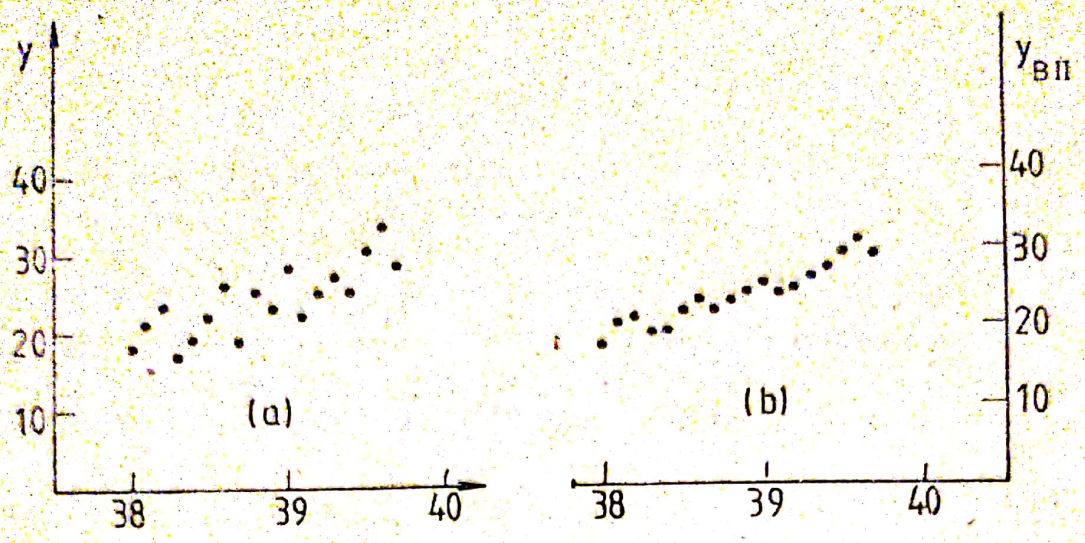


Fig. 5.3. **Exemplu de netezire a unei curbe empirice.**

zătoare este reprezentată grafic în figura 5.3a, de unde se vede că există o dispersie destul de mare.

Aplicînd formulele (5.23) și (5.33) rezultă ordonatele y_{AI} și respectiv y_{BI}, după prima netezire. Apoi repetînd procedeul de netezire, adică considerînd acum mărimile y_{AI} și y_{BI} ca și coordonate experimentale, aplicăm a doua oară formulele (5.23) și (5.33) și rezultă ordonatele y_{AII} și y_{BII}. În figura 5.3b am reprezentat curba empirică netezită de două ori cu formula (5.33).

Desigur, și în figura 5.3b mai există o oarecare dispersie, totuși sub acest aspect curba respectivă poate fi utilizată în scopuri practice.

b) *Cazul curbei cu curbură mare*

Pentru motivele arătate mai înainte, și aici vom prezenta două procedee de „netezire" a unei curbe empirice: i) procedeul bazat pe metoda celor mai mici pătrate, și ii) procedeul care se bazează pe valoarea medie a 10 parabole care se pot duce prin cele 10 grupe de cîte trei puncte distincte care se pot forma dintr-o serie de cinci puncte observate consecutiv.

i) În cazul cînd curba empirică prezintă o curbură destul de mare și nu se poate utiliza ipoteza din paragraful precedent, adică nu se poate considera că prin fiecare trei puncte consecutive trece o linie dreaptă, atunci pe un interval destul de mic vom aproxima funcția respectivă printr-o parabolă care să treacă prin cel puțin cinci puncte consecutive. În acest caz putem scrie

$$y = a_i x^2 + b_i x + c_i \tag{5.35}$$

Scanned with OKEN Scanner

Urmărind raționamentul din paragraful precedent, cele mai probabile valori a_i, b_i și c_i pentru secvența i, se vor obține din condiția de minim a funcției

$$f(a_i, b_i, c_i) = \sum_{j=1}^{5} (a_i x_{ij}^2 + b_i x_{ij} + c_i - y_{ij})^2 \tag{5.36}$$

Prin urmare, anulînd derivatele parțiale de ordinul întîi în raport cu a_i, b_i și c_i, avem

$$\sum_{j=1}^{5} (a_i x_{ij}^2 + b_i x_{ij} + c_i - y_{ij}) = 0$$

$$\sum_{j=1}^{5} (a_i x_{ij}^2 + b_i x_{ij} + c_i - y_{ij}) x_{ij} = 0$$

$$\sum_{j=1}^{5} (a_i x_{ij}^2 + b_i x_{ij} + c_i - y_{ij}) x_{ij}^2 = 0$$

sau

$$\left.\begin{aligned} 5c_i \quad + b_i \Sigma x_{ij} + a_i \Sigma x_{ij}^2 &= \Sigma y_{ij} \\ c_i \Sigma x_{ij} + b_i \Sigma x_{ij}^2 + a_i \Sigma x_{ij}^3 &= \Sigma y_{ij} x_{ij} \\ c_i \Sigma x_{ij}^2 + b_i \Sigma x_{ij}^3 + a_i \Sigma x_{ij}^4 &= \Sigma y_{ij} x_{ij}^2 \end{aligned}\right\} \tag{5.37}$$

unde i reprezintă numărul de ordine al unei secvențe de cinci puncte utilizate, iar j indică numărul de ordine al punctelor din cadrul aceleiași secvențe: $j = \overline{1;5}$.

Avînd în vedere faptul că mărimile x_{ij}, y_{ij} sînt cunoscute din activitatea experimentală, relațiile (5.37) definesc un sistem de trei ecuații cu cele trei necunoscute: a_i, b_i și c_i. În modul acesta, pentru fiecare interval i, putem scrie

$$y_{i+2} = a_i x_{i+2}^2 + b_i x_{i+2} + c_i \tag{5.38}$$

unde, în cazul prezent, $i + 2$ definește punctul de la mijlocul unei secvențe de cinci puncte utilizate. Pentru prima secvență avem: $i = 1$, $i + 2 = 3$ etc, adică este vorba de corectarea ordonatei punctului al treilea. În cazul intervalului al doilea: $i = 2$, iar o formulă analoagă cu (5.38) ne va permite să corectăm ordonata y_4 etc.

Avînd în vedere faptul că datele experimentale pot fi destul de numeroase, sistemul (5.37) va trebui să fie programat pentru un calculator electronic.

În cazul cînd punctele observate au abscisele echidistante, adică atunci cînd $h = x_{i+1} - x_i = \text{constant}$, rezolvarea sistemului (5.37) se simplifică în mod considerabil.

Scanned with OKEN Scanner

Dacă pentru măsurarea absciselor celor cinci puncte dintr-o anumită secvență alegem punctul din mijloc ($j = 3$) ca origine, atunci

$$x_1 = -2h,\quad x_2 = -h,\quad x_3 = 0,\quad x_4 = h,\quad x_5 = 2h,$$

unde $j = \overline{-2\,;\,+2}$. În aceste condiții, pentru sumele din sistemul (5.37) obținem:

$$\left.\begin{aligned}
\sum_{j=-2}^{+2} x_{ij} &= -2h - h + h + 2h = 0\\
\sum_{j=-2}^{+2} x_{ji}^2 &= 4h^2 + h^2 + h^2 + 4h^2 = 10h^2\\
\sum_{j=-2}^{+2} x_{ij}^3 &= -8h^3 - h^3 + h^3 + 8h^3 = 0\\
\sum_{j=-2}^{+2} x_{ij}^4 &= 16h^4 + h^4 + h^4 + 16h^4 = 34h_4
\end{aligned}\right\} \qquad (5.39)$$

iar relația (5.38) devine

$$\tilde{y}_{i+2} = c_i \qquad (5.40)$$

După cum se vede, ordonata corectată va fi egală cu mărimea c_i pentru determinarea căreia, ținînd seama de relațiile (5.39), din prima și ultima ecuație a sistemului (5.37), avem

$$\left.\begin{aligned}
5c_i + 10h^2a_i &= y_{i-2} + y_{i-1} + y_i + y_{i+1} + y_{i+2}\\
10c_ih^2 + 34h^4a_i &= (4y_{i-2} + y_{i-1} + y_{i+1} + 4y_{i+2})h^2
\end{aligned}\right\} \qquad (5.41)$$

Dacă înmulțim prima ecuație cu $+34h^2$ iar a doua cu -10 și adunăm rezultatele respective, avem

$$\tilde{y}_{i+2} = c_i = \frac{-3y_{i-2} + 12y_{i-1} + 17y_i + 12y_{i+1} - 3y_{i+2}}{35} \qquad (5.42)$$

Dacă, în partea dreaptă a formulei (5.42), adunăm și scădem mărimea y_i, avem mai departe

$$\tilde{y}_i = y_i - \frac{3}{35}(y_{i-2} - 4y_{i-1} + 6y_i - 4y_{i+1} + y_{i+2}) \qquad (5.43)$$

sau

$$\tilde{y}_i = y_i - \frac{3}{35}\Delta^{IV} y_{i-2} \qquad (5.44)$$

unde $\Delta^{IV} y_{i-2}$ este diferența de ordinul patru a funcției $y = f(x)$ pentru $x \in (x_{i-2},\ x_{i+2})$.

Scanned with OKEN Scanner

Aplicație numerică

Să considerăm datele experimentale din tabela 5.3

Tabela 5.3

x	-10	-9	-8	-7	-6
y	8	2	1	2	8

Efectuînd înlocuirile corespunzătoare în formula (5.43) obținem ordonata care corespunde abscisei $x = -8$,

$$y(-8) = 1 - 0{,}0857\,(8 - 8 + 6 - 8 + 8) = 1{,}0 - 0{,}51 = 0{,}49$$

adică, curba netezită va trebui să treacă prin punctul de coordonate $x = -8$ și $y = 0{,}49$, și nu prin punctul observat de coordonate $x = -8$ și $y = 1$.

În modul prezentat aici putem corecta ordonatele tuturor punctelor dintr-o serie de date cu excepția primelor și ultimelor două puncte care vor rămîne nemodificate.

ii) În vederea prezentării procedeului care se bazează pe valoarea medie a ordonatei determinată de cele 10 parabole care pot fi trasate prin cele 10 grupe de cîte trei puncte din secvența celor 5 puncte succesive observate, vom utiliza formula generală care reprezintă ecuația unei astfel de parabole. Dacă p_a, p_b și p_c ar fi trei puncte dintr-o secvență considerată, atunci parabola generală va avea expresia (vezi Todoran, 1976, p. 159)

$$y = y_b + \frac{x - x_b}{x_a - x_c}\left[\frac{y_a - y_b}{x_a - x_b}(x - x_c) - \frac{y_c - y_b}{x_c - x_b}(x - x_a)\right] \qquad (5.45)$$

Scriind 10 formule de forma (5.45) pentru $x = x_3$, adică pentru punctul de la mijlocul secvenței considerate și adunîndu-le, obținem

$$\begin{aligned} 10\tilde{y}_3 = {} & 4y_3 + 3y_2 + 3y_4 + \frac{x_3 - x_2}{x_3 - x_1} \cdot \frac{y_3 - y_2}{x_3 - x_2}(x_3 - x_1) + \\ & + \frac{x_3 - x_2}{x_1 - x_4}\left[\frac{y_1 - y_2}{x_1 - x_2}(x_3 - x_4) - \frac{y_4 - y_2}{x_4 - x_2}(x_3 - x_1)\right] + \\ & + \frac{x_3 - x_2}{x_1 - x_5}\left[\frac{y_1 - y_2}{x_1 - x_2}(x_3 - x_5) - \frac{y_5 - y_2}{x_5 - x_2}(x_3 - x_1)\right] + \qquad (5.46) \\ & + \frac{x_3 - x_4}{x_3 - x_5}\frac{y_3 - y_4}{x_3 - x_4}(x_3 - x_5) + \frac{x_3 - x_4}{x_1 - x_5}\left[\frac{y_1 - y_4}{x_1 - x_4}(x_3 - x_5) - \frac{y_5 - y_4}{x_5 - x_4}(x_3 - x_1)\right] + \\ & + \frac{x_3 - x_4}{x_2 - x_5}\left[\frac{y_2 - y_4}{x_2 - x_4}(x_3 - x_5) - \frac{y_5 - y_4}{x_5 - x_4}(x_3 - x_2)\right]. \end{aligned}$$

Scanned with OKEN Scanner

Prin urmare, dacă dispunem de o secvență de cinci puncte observate, pentru corectarea („netezirea") ordonatei punctului de la mijloc se fac înlocuirile în formula (5.46), se împart la 10 și avem mărimea $\tilde{y}_3$.

Formula (5.46) se poate utiliza în cazul general cînd pasul absciselor este variabil. Dar dacă dispunem de serii de observații cu pasul absciselor constant, adică dacă $h = x_{i+1} - x_i = \text{constant}$, atunci formula (5.46) devine

$$\tilde{y}_3 = \frac{-4y_1 + 16y_2 + 36y_3 + 16y_4 - 4y_5}{60}$$

unde simplificăm cu 4 și considerăm o secvență generală de cinci puncte ($i = 3$) astfel încît

$$\tilde{y}_i = \frac{-y_{i-2} + 4y_{i-1} + 9y_i + 4y_{i+1} - y_{i+2}}{15} \qquad (5.47)$$

Aplicația practică a formulei (5.47) am efectuat-o tot la datele din tabela 5.3 astfel că

$$\tilde{y}_{-8} = \frac{-8 + 4.2 + 9 + 4.2 - 8}{15} = \frac{9}{15} = +0{,}60.$$

Prin urmare, atît din comparația formulelor (5.47) și (5.42) cît și din aplicațiile practice ale acestora se vede că ambele formule reprezintă medii ponderate, deosebirea dintre ele fiind aceea că în formula (5.47) ordonata care urmează a fi „netezită" intră cu o pondere mai mare, astfel că „netezirea" devine mai puțin dură.

5.4. Deviația standard a unei curbe empirice

Noțiunea de deviație standard a fost întîlnită în capitolul III [vezi formula (3.5)] unde ne-am referit la media aritmetică a mai multor date experimentale. În capitolul prezent ne propunem să arătăm modul în care, prin intermediul acestei noțiuni, putem să atribuim un anumit calificativ comparativ unor curbe empirice. Vom urmări această problemă la cele două cazuri prezentate mai sus pentru trasarea curbelor empirice: a) trasarea curbelor prin puncte medii, și b) ajustarea datelor experimentale.

a) *Deviația standard a unei curbe medii*

Am văzut mai înainte că atunci cînd datele experimentale sînt grupate în puncte medii — puncte normale, se obțin valorile medii atît ale absciselor, x, cît și ale ordonatelor, y. Diferența dintre valorile individuale (datele observate) și mărimile medii ne permit, prin inter-

mediul formulei (3.5), să calculăm deviaţiile standard corespunzătoare, adică

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{m}}, \qquad \sigma_y = \sqrt{\frac{\sum_{i=1}^{m}(y_i - \bar{y})^2}{m}} \tag{5.48}$$

unde sumele de sub radicali se referă la toate cele m puncte observate din care s-a alcătuit un punct normal. În aceste condiţii, deviaţia standard a unui punct mediu va fi definită prin formula

$$\sigma_j^2 = (\sigma_x^2 + \sigma_y^2)_j \tag{5.49}$$

unde indicele j arată numărul de ordine al unui punct din secvenţa de puncte normale calculate.

Pentru a evalua gradul general de precizie al unei curbe empirice, putem calcula deviaţia standard, σ, astfel încît să avem

$$\sigma = \sqrt{\frac{\sum_{j=1}^{n}(\sigma_j - \bar{\sigma})^2}{n}} \tag{5.50}$$

unde n reprezintă numărul total de puncte normale existente, iar $\bar{\sigma}$ este deviaţia standard medie a deviaţiilor individuale ale punctelor normale.

Formula (5.50) defineşte un parametru care dă un fel de calificativ pentru curba empirică discutată. Acest parametru se poate calcula pentru mai multe curbe empirice şi constituie astfel un mijloc de a atribui anumite ponderi din care să rezulte şi gradul de probabilitate a corelaţiei şi prognozei care se deduc din astfel de curbe.

b) *Deviaţia standard a unei curbe „netezite"*

În cazul cînd pentru trasarea unei curbe empirice ne-am folosit de metodele de ajustare a datelor experimentale, deviaţia standard se poate evalua într-o formă asemănătoare cu aceea prezentată mai sus, dar cu mici modificări.

După cum am văzut mai înainte, prin procedeele de ajustare a datelor experimentale, abscisele rămîn nemodificate, iar pentru ordonate se aplică anumite corecţii. Diferenţele dintre ordonatele observate, y, şi cele corectate, $\tilde{y}$, sînt singurele mărimi care pot fi utilizate pentru evaluarea gradului relativ de precizie a unei curbe empirice. În acest scop, dacă n este numărul total de puncte observate, deviaţia standard a unei curbe „netezite" se poate defini prin relaţia

$$\sigma = \sqrt{\frac{\sum_{k=1}^{n}(y_k - \bar{y}_k)^2}{n}} \tag{5.51}$$

Scanned with OKEN Scanner

unde y_k este ordonata observată a punctului pentru care $n = k$, iar $\tilde{y}_k$ este ordonata corectată prin intermediul uneia din formulele: (5.19), (5.23), (5.32), (5.33), (5.38), (5.42), (5.46) sau (5.47).

Observație. Am amintit aici că deviația standard poate constitui un parametru necesar pentru atribuirea unei ponderi relative la mai multe curbe empirice examinate. Trebuie, însă, precizat că astfel de probleme se pot aborda numai la curbele empirice care au fost supuse aceluiași fel de prelucrare matematică: curbe normale (medii) sau curbe empirice la care s-a aplicat un anumit procedeu de netezire.

5.5. Rolul simetriei curbelor empirice

Unele dintre problemele amintite mai înainte — utilizarea curbelor empirice în probleme de prognoză — se simplifică considerabil dacă astfel de curbe prezintă anumite particularități cum sînt, de exemplu, simetria sau periodicitatea.

Să considerăm că avem de a face cu un fenomen al cărui mod de manifestare duce la o curbă simetrică. Dacă din activitatea practică dispunem de un arc mic din porțiunea simetrică a acesteia, pe partea simetrică lui, putem trasa curba cu destul de mare încredere chiar și acolo unde nu dispunem de date experimentale. Într-un astfel de caz extrapolarea este mult mai precisă și, prin aceasta, se va obține o prognoză mai bună.

În figura 5.4 am reprezentat curba de variație a strălucirii unei stele variabile. Este vorba de eclipsarea reciprocă a componentelor unei stele duble fotometrice. Din modul de desfășurare a unor astfel de fenomene se știe că în vecinătatea minimului principal, la abscisa

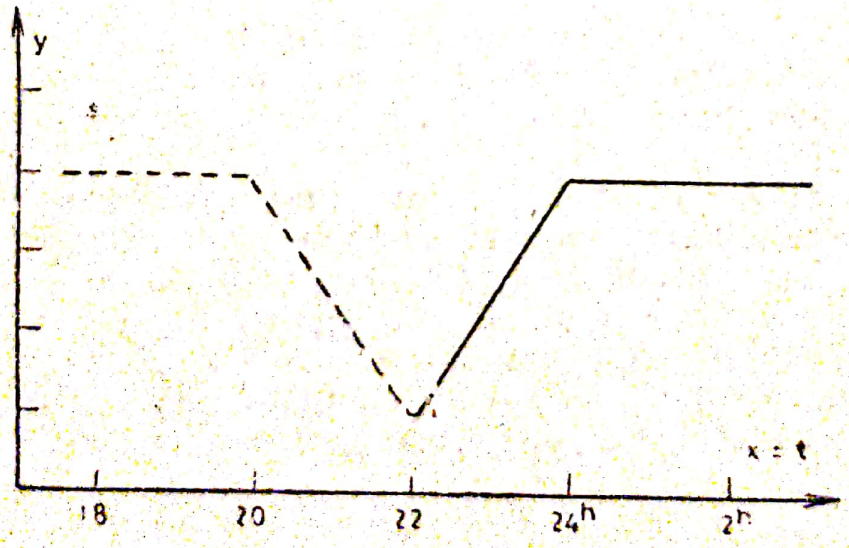


Fig. 5.4. Curba de variație a strălucirii unei stele duble fotometrice (Stea dublă cu eclipse).

cînd strălucirea este minimă, curba de lumină este simetrică. Prin urmare, cu o probabilitate destul de mare, putem prevedea sfîrșitul eclipsei observate. În figura 5.4 linia punctată reprezintă curba observată de unde stabilim faptul că eclipsa a început la orele 20 (cînd începe să scadă strălucirea), are faza maximă la orele 22 cînd avem cea mai mică strălucire și apoi, prin linii continue, am schițat curba prin extrapolare. De aici rezultă că eclipsa se va termina la orele 24, această prognoză putînd fi verificată și confirmată prin observații.

În cadrul datelor experimentale, de obicei simetria se pune în evidență față de o dreaptă paralelă cu una din axele de coordonate ale sistemului în care au fost obținute datele respective.

Desigur, constatarea prezenței unei simetrii în cadrul curbelor empirice duce la postularea mult mai sigură a cauzalității. Astfel, de exemplu, simetria curbei de lumină la un sistem dublu de stele strînse arată că cele două componente se mișcă pe orbite circulare în jurul centrului de masă, fapt de mare importanță pentru problemele teoretice privind determinarea parametrilor fizici ai stelelor componente.

Matematic, asimetria se exprimă prin (vezi Marinescu et al, 1966, p. 109).

$$A_s = \left(\frac{\mu_3^2}{\mu_2^3}\right)^{1/2} = \frac{\mu_3}{\mu_2^{3/2}} \tag{5.52}$$

unde

$$\mu_k = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^k n_i}{n} \tag{5.53}$$

se numesc momente de ordinul k.

Dacă scriem formula (5.53) pentru $k = 2$, și 3, după efectuarea înlocuirilor în (5.52) avem:

$$A_s = \frac{\sqrt{n} \sum_{i=1}^{n} (x_i - \bar{x})^3 n_i}{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2 n_i\right]^{3/2}} \tag{5.54}$$

unde n este numărul total de puncte empirice, iar n_i reprezintă numărul de puncte $M_i(x_i, y_i)$.

Dacă, atît pe ramura ascendentă cît și pe ramura descendentă putem dispune de puncte care să aibă aceeași ordonată, asimetria

Scanned with OKEN Scanner

curbei respective se poate examina și prin analiza valorilor obținute pentru abscisele medii ale perechilor de astfel de puncte. Prin urmare dacă x_i și x_i' sînt abscisele celor două puncte care au ordonata y_i, dar sînt situate pe ramuri diferite, abscisa medie este definită prin

$$\bar{x}_i = \frac{x_i + x_i'}{2} \tag{5.55}$$

În cazul cînd o anumită curbă empirică este asimetrică, valorile $\bar{x}_i$ definite prin relația (5.55) vor prezenta o dependență funcțională. La curbele simetrice dreapta $x = \bar{x}_i$ va fi axă de simetrie.

Aplicație practică

În primele două coloane ale tabelei 5.4 se găsesc coordonatele unei serii de 24 de puncte empirice, cîte 12 pe o ramură.

Tabela 5.4

x	y	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^3$	$\frac{x+x'}{2}$
25	30	−42	1764	−74088	87,5
26	35	−41	1681	−68921	85 5
27	40	−40	1600	−64000	81,0
28	45	−39	1521	−59319	79,0
29	50	−38	1444	−54827	74,5
30	55	−37	1369	−50653	70,0
31	60	−36	1296	−46656	64,0
32	65	−35	1225	−42875	61,0
33	70	−34	1156	−39304	56,5
34	75	−33	1089	−35973	52,0
37	80	−30	900	−27000	48,5
40	83	−27	729	−19683	45,0
50	83	−17	289	− 4913	
60	80	− 7	49	− 343	
70	75	3	9	27	
80	70	13	169	2197	
90	65	23	529	12167	
97	60	30	900	27000	
110	55	43	1849	79507	
120	50	53	2809	148877	
130	45	63	3969	250047	
135	40	68	4624	314432	
145	35	78	6084	474552	
150	30	83	6889	571787	
			43943	1292029	

Scanned with OKEN Scanner

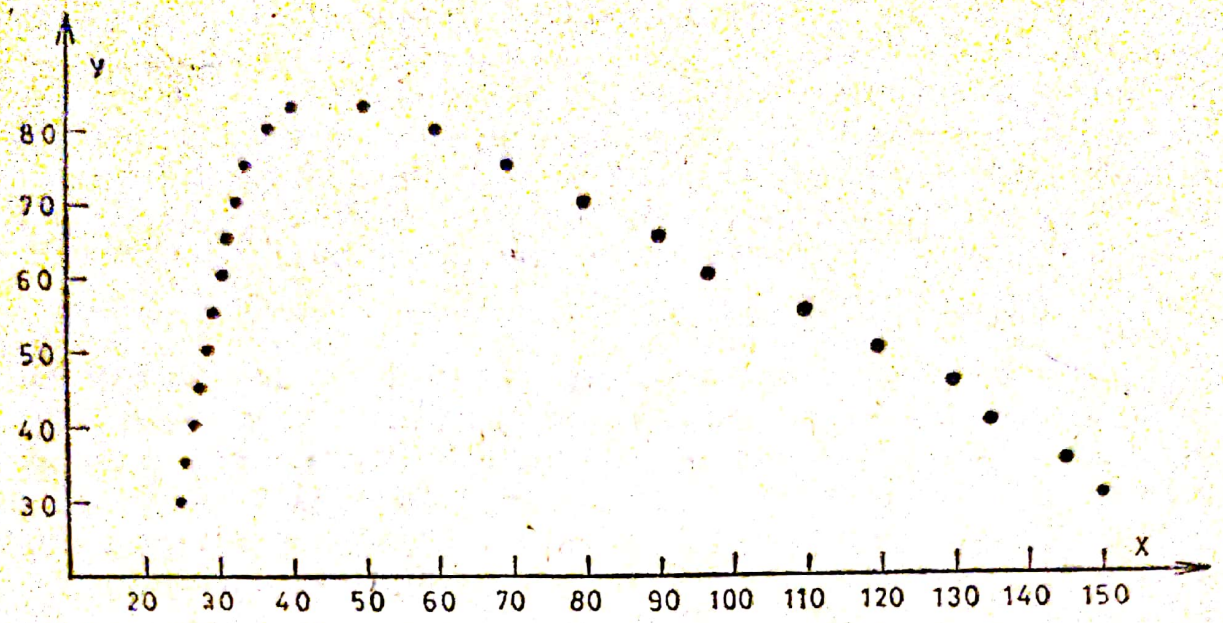


Fig. 5.5. Exemplu de curbă asimetrică.

Deoarece aici am avut $n_i = 1$, la coloanele 4 și 5 am avut o însumare obișnuită. Prin urmare:

$$n = 24, \qquad \sqrt{n} = 4{,}898979, \qquad \bar{x} = \frac{1609}{24}, = 67$$

$$(x_i - \bar{x})^2 = 43943, \qquad (x_i - \bar{x})^3 = 1292029, \qquad 43943^{1/2} = 210,$$

$$A_s = \frac{4{,}898979 \times 1292029}{43943 \times 210} = \frac{6329623}{9211589} = 0{,}687$$

În ultima coloană a tabelei 5.4 avem abscisele medii ale punctelor care au aceeași ordonată. Și aici asimetria este evidentă. De altfel, graficul din figura 5.5 arată imediat că este vorba de o curbă asimetrică la care ramura ascendentă este mult mai abruptă decît ramura descendentă.

5.6. Proiecte de prognoză bazate pe corelație

Prognoza, privind manifestarea anumitor fenomene sau existența unor obiecte încă nedetectate, trebuie să se bazeze pe un număr foarte mare de date care să fie interpretate corect din punct de vedere matematic. Iată de ce, în cadrul expunerii de pînă acum, am ilustrat doar modul de aplicare practică a unor procedee care constituie *părți componente* ale unei prognoze. Pînă acum am privit prognoza sub aspectul său mai mult sau mai puțin general, pe cînd problemele specifice ale unor domenii concrete au rămas să fie detaliate de către specialiștii respectivi.

Scanned with OKEN Scanner

În prezent noi nu avem date numerice suficiente pentru a elabora o anumită prognoză, dar nu avem nici spațiu necesar pentru așa ceva. În schimb, în urma celor expuse mai înainte, putem să discutăm anumite scheme de principiu, scheme care să poată fi ușor aplicate de către specialiștii care au acces la datele empirice necesare. De aceea, ne-am propus să trecem în revistă cîteva proiecte de prognoză pentru unele dintre cele mai enigmatice situații.

a) Curbele medii anuale de temperatură

Să considerăm că, de la Institutul Meteorologic, dispunem de date suficiente pentru a întocmi puncte medii pe zile, privind variația temperaturii aerului, T, în funcție de timpul t. În acest caz, întocmind grafice empirice ale funcției $T = f(t)$ pe o secvență mare de ani, se poate evidenția, bineînțeles dacă există, vreo anomalie remarcabilă cum ar fi, de exemplu, scăderea bruscă a temperaturii în jurul datei de 11 mai (la „Sfinții de Gheață").

Unele anomalii astfel detectate ar putea arăta dacă este vorba de o cauză temporară care se manifestă numai într-un an sau există cauze care acționează în fiecare an cam la aceeași dată (ploile de vară din jurul datei de 20 iulie).

În anumite cazuri s-ar putea detecta cauzalități periodice care se manifestă după un anumit număr de ani. În toate aceste cazuri o analiză detaliată a curbelor empirice permite meteorologilor să elaboreze prognoze destul de valoroase. Astfel de probleme sînt bine cunoscute, dar noi le-am reprodus aici pentru a da exemple de corelații pentru seriile dinamice la care faza de corelație ar putea aduce informații interesante.

b) Tratamentul chirurgical și fazele Lunii

Se vorbește că există medici chirurgi care preferă să efectueze operațiile la anumite faze ale Lunii. Nu se cunosc asemenea cercetări care să pună în evidență o astfel de relație. În orice caz, ar fi deosebit de interesant și util dacă s-ar verifica existența unei corelații între gradul de reușită a operațiilor și pozițiile Lunii pe bolta cerească.

Este adevărat că nimeni nu cunoaște mecanismul prin care Luna ar putea influența operațiile chirurgicale, dar se cunosc multe evenimente în care Luna are influență asupra organismului uman mai ales asupra celui feminin.

Se bănuiește că unele instabilități meteorologice ar fi legate de fazele Lunii, dar un studiu detaliat bazat pe corelație încă nu există. Deatlfel, persoanele reumatologice sînt sensibile la schimbarea vremii și prin aceasta, pot fi sensibile și la modificarea pozițiilor Lunii pe bolta cerească.

Desigur, dacă s-ar dispune de o curbă empirică privind gradul de reușită a operațiilor chirurgicale în funcție de fazele Lunii, opera-

Scanned with OKEN Scanner

ţiile care suferă aminări ar putea fi programate pentru epocile cele mai favorabile. Astfel de grafice s-ar putea întocmi cu uşurinţă după scriptele existente în anumite clinici profilate pe astfel de tratamente, pentru gradul de uşurinţă a operaţiilor putînd fi stabilite trepte numerice.

c) *Cometele şi iernile geroase*

Există persoane care, în mod mai mult sau mai puţin conştient, leagă iarna geroasă a anului 1985 de apropierea cometei Halley, astfel de legături însă nu sînt cercetate destul de bine din punct de vedere statistic. Iată de ce, cercetarea unei corelaţii dintre apariţiile cometelor şi iernile geroase ar fi deosebit de interesantă din mai multe puncte de vedere.

De altfel, au mai existat evenimente sociale care erau „puse pe seama unor comete", dar cercetările statistice şi fizice au arătat că nu era vorba de corelaţii ferme. Aici, un rol l-a jucat şi imaginaţia omului şi mai ales necunoaşterea adevăratei naturi a aştrilor cu coadă. (Există multe cazuri descrise în literatură în care se încearcă să se lege războaiele şi căderile unor capete mari de apariţia unor comete celebre).

d) *Catastrofele sociale şi corelaţia*

Răsfoind literatura existentă, şi selecţionînd momentele unor catastrofe sociale: războaie, epidemii, cutremure, erupţii vulcanice, inundaţii, secete etc. s-ar putea aduna datele necesare pentru întocmirea tabelelor de corelaţie pentru studiul lagăturii dintre acestea şi unii factori cosmici. Evident, rezultatele obţinute vor diferi de la o epocă la alta şi din cauza unor acţiuni sociale care au fost întreprinse la epoci diferite: lupta pentru pace, impunerea unor măsuri sanitare, rolul unor factori educativi etc.

Nimeni nu a vorbit pînă în prezent despre influenţa pe care ar putea s-o aibă exploziile grandioase din nucleele active ale unor galaxii, nimeni nu s-a ocupat de rolul exploziilor de nove şi supernove etc.

Am amintit aici doar cîteva dintre problemele în care teoria corelaţiei ar putea aduce rezultate importante, dar pentru ca astfel de cercetări să fie eficiente, trebuie adunate mari cantităţi de date empirice. Cele expuse pînă acum ar putea fi un îndemn pentru activităţi de acest fel.

Scanned with OKEN Scanner

CAPITOLUL VI

PROGNOZA FENOMENELOR PERIODICE

În cadrul problemelor de prognoză un loc deosebit îl ocupă fenomenele pentru care s-a stabilit deja caracterul lor periodic. În astfel de cazuri, în mod deosebit ne interesează intervalul de timp — perioada — după care un anumit femomen se va manifesta într-un mod aproximativ cunoscut de noi. Se pot da multe exemple de acest fel: timpul după care, într-un anumit loc, un cutremur de Pămînt se va repeta cu aceeaşi intensitate; timpul dintre două erupţii ale unui vulcan; timpul dintre două inundaţii consecutive; timpul dintre două sau mai multe epidemii etc.

Deşi unele dintre perioadele menţionate mai sus pot fi aparente, pentru problemele de prognoză ele devin deosebit de importante. Astfel, de exemplu, dacă ne referim la erupţiile unui vulcan, activitatea vulcanului deşi ar putea să fie periodică, cîteva înregistrări nu ne pot spune mare lucru despre activitatea vulcanului deoarece aceste izbucniri ar putea face parte dintr-un mod mult mai general de comportare periodică. O adevărată periodicitate, care să ducă la posibilitatea postulării unor condiţii fizice determinante, se poate stabili numai cînd dispunem de observaţii „uniform" distribuite pe un interval mare de timp. Astfel, de exemplu, pot fi amintite eclipsele de Lună şi de Soare care au fost observate de-a lungul mai multor sute de ani. De aceea, deşi natura lor nu era cunoscută, astronomii din antichitate le prevedeau cu o precizie destul de bună.

De asemenea, se cunoşteau periodicităţile privind revărsările unor rîuri, dar acestea erau ţinute în secret faţă de marele public.

În cadrul fenomenelor periodice bine studiate s-au putut deosebi trei grupe, după cum perioadele respective erau simple şi constante, simple dar variabile, şi grupa fenomenelor care erau determinate de prezenţa mai multor perioade. Fiecare dintre aceste trei grupe de fenomene îşi are caracteristicile sale specifice, fapt care ne determină să le analizăm separat.

6.1. Caracteristicile principale ale unei funcţii periodice

În cazul curbelor empirice periodice există cîţiva parametri a căror explicaţie este necesară.

a) *Perioada.*

O funcţie reală definită pe mulţimea numerelor reale $f : R \to R$,

se numește periodică, de perioadă P, dacă există un număr $P \neq 0$ astfel încît pentru orice x din R ($x \in R$) să avem relația

$$f(x + P) = f(x). \tag{6.1}$$

Din această definiție rezultă că, dacă P este o perioadă, pentru orice număr întreg E numărul real $E.P$ este de asemenea o perioadă.

Prin urmare, perioada apare ca un interval de timp după care se repetă funcția $f(x)$. Aceasta este un parametru important în probleme de prognoză și contribuie direct la îmbogățirea cunoștințelor noastre despre lumea înconjurătoare.

Așadar, o funcție empirică este periodică dacă ea îndeplinește condiția (6.1). Totuși trebuie să specificăm aici și faptul că functia empirică se poate considera periodică și atunci cînd valorile ei se repetă în cadrul limitei preciziei de determinare a ordonatelor corespunzătoare.

Datorită erorilor accidentale s-ar putea ca ordonatele care corespund la abscisele x_i și $x_i + P$ să fie diferite, dar această diferență trebuie să fie mai mică decît dispersia datelor experimentale.

Așa cum am subliniat și mai sus, perioada este intervalul de timp după care se repetă forma curbei empirice, fapt care ne atrage atenția asupra unor detalii ale acestor curbe. Este vorba de a urmări pozițiile unor puncte caracteristice cum sînt, de exemplu, punctele de maxim și minim sau punctele care determină anumite valuri de-a lungul undei generale. În anumite situații practice am putea defini perioada ca intervalul de timp dintre abscisele a două maxime sau minime consecutive. Pozițiile unor astfel de puncte se determină relativ ușor, fapt care contribuie la determinarea precisă a perioadei.

În figura 6.1 am reprezentat schematic o funcție periodică $y=f(x)$ în vederea precizării celor mai importante caracteristici. Aici

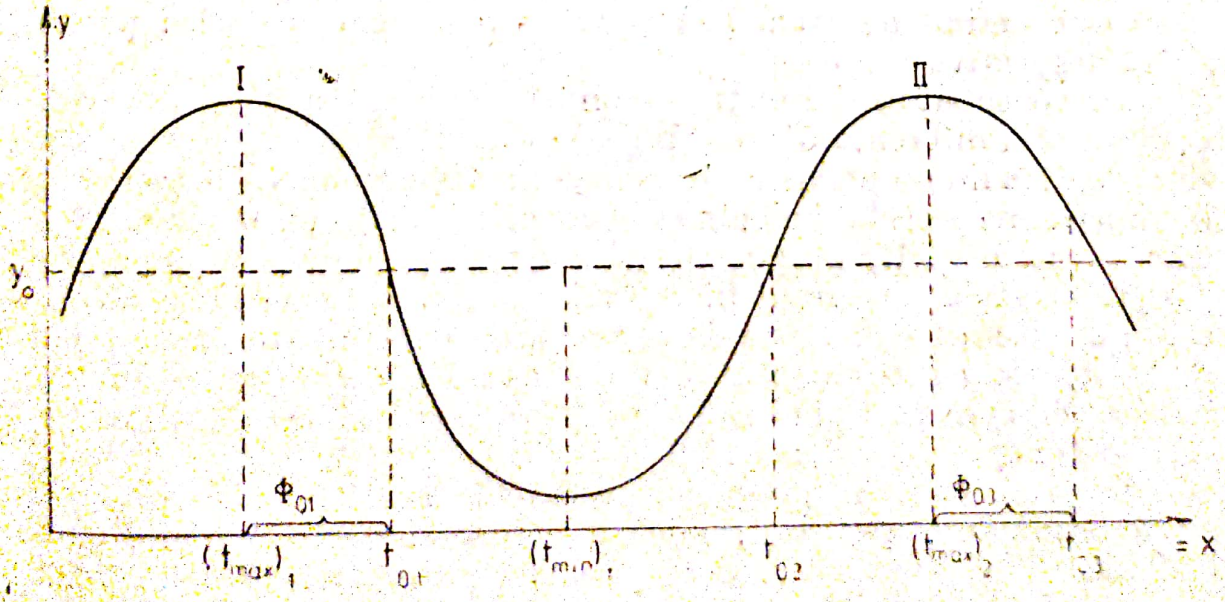


Fig. 6.1. Definirea fazei unei funcții periodice.

$(t_{max})_2$ și $(t_{max})_1$ reprezintă abscisele a două maxime succesive. În acest caz perioada se poate defini prin relația

$$P = (t_{max})_2 - (t_{max})_1 \tag{6.2}$$

Valorile cele mai mari sau cele mai mici, pe care le poate avea o funcție empirică, definesc punctele de extremum global sau principal. Dar, pe ciclul unei funcții periodice, uneori, putem distinge și puncte de extremum cu amplitudini mai mici decît cele ale extremelor globale, este vorba de extreme locale sau secundare.

Curba $y = f(x)$ poate să fie simetrică sau asimetrică, în ambele cazuri relația (6.2) își păstrează valabilitatea. Totuși, așa cum vom vedea mai departe, la o curbă simetrică abscisele punctelor de extremum se determină mai ușor și, prin aceasta, durata determinată a perioadei va fi mult mai precisă.

Dacă $2t_{min} \neq (t_{max})_2 + (t_{max})_1$. curba este asimetrică în timp. Asimetria în ordonată se tratează prin metoda discutată în capitolul precedent.

Pe curba din figura 6.1 am ales ordonata y_0 și am trasat (punctat) dreapta $y = y_0$. De aici se vede că această ordonată corespunde atît la puncte de pe ramuri ascendente, cît și la puncte de pe ramurile descendente. Cu ajutorul unei astfel de ordonată de asemenea putem defini perioada unei funcții empirice periodice. *Se numește perioadă, P, intervalul de timp scurs între abscisele a două puncte de aceeași ordonată și de pe același fel de ramură* (ascendente sau descendente și consecutive).

b) *Ciclul unei curbe empirice*

Porțiunea dintr-o curbă periodică, care corespunde diferenței de abscise cu o perioadă se numește *ciclu*. Pe figura 6.1 un astfel de ciclu este definit de arcul $\overset{\frown}{\text{I.II}}$ și măsoară lungimea curbei pe intervalul respectiv.

Să considerăm o funcție empirică periodică a cărei perioadă o definim prin diferența dintre abscisele a două extreme globale succesive. Deși, în limita preciziei cu care s-au obținut datele experimentale, funcția respectivă îndeplinește condiția (6.1), pe o anumită porțiune a unui ciclu se poate sesiza prezența unui extremum local, poziția acestuia din urmă fiind variabilă de la un ciclu la altul. În astfel de situații este vorba de niște variații secundare care ar putea să fie și ele periodice, s-ar detecta astfel o perioadă „secundară", mărime de care va trebui să se țină seamă la definirea precisă a perioadei funcției considerate.

În problemele de prognoză pe care le urmărim aici considerăm că determinarea pe cale grafică a perioadei poate fi mulțumitoare mai ales dacă avem în vedere procedeele de interpolare și netezire ale curbelor empirice. Pentru determinarea mult mai precisă și o

Scanned with OKEN Scanner

teorie mult mai matematizată recomandăm cercetările publicate cu ani în urmă (vezi Todoran, 1967, paragrafele 9.4, 9.5 și capitolul 10 în întregime). Determinarea perioadei prin metoda corelației am expus-o în paragraful 3.11 b.

c) *Efemerida unei funcții periodice*

Fie t_0 momentul (abscisa) primului maxim observat, sau al unui alt punct caracteristic de pe curba empirică.

Dacă presupunem cunoscută perioada P, al doilea maxim observat, primul după maximul admis ca origine pentru măsurarea timpului, va avea abscisa t_1, la al doilea va corespunde t_2 etc. În aceste condiții putem scrie:

$$\begin{aligned} t_1 &= t_0 + P \\ t_2 &= t_1 + P = t_0 + 2P \\ &\cdots\cdots\cdots\cdots \\ t_E &= t_0 + E.P \end{aligned} \qquad (6.3)$$

unde E reprezintă un număr întreg de perioade sau epoci.

Formula (6.3) permite să se obțină o serie de valori t_E pentru seria corespunzătoare de numere întregi E, adică ea va exprima o serie de valori numerice dintre care unele vor corespunde timpului viitor. Din acest motiv, acestei funcții i-am atribuit denumirea de efemeridă și reprezintă modul cel mai elegant posibil pentru elaborarea unei prognoze.

Expresia (6.3) a fost introdusă de către observatorii de stele variabile care, prin intermediul ei, își planificau activitatea observațională astfel încît să poată urmări atît ramura crescătoare cît și cea descrescătoare. Cu toate acestea, formula respectivă exprimă un mod general de a elabora o prognoză.

Avînd în vedere faptul că în formula (6.3) momentul t_E este o funcție de numărul întreg E la puterea întîia, această expresie a primit denumirea de *efemeridă liniară.*

d) *Faza ca unitate de măsură a timpului în interiorul unui ciclu*

Dacă perioada unei funcții este admisă ca unitate de timp, în interiorul acesteia se pot introduce unități mai mici, dar utile pentru scopul urmărit.

Astfel, de exemplu, pe fig.6.1 intervalul de timp $\Phi_{01}=t_{01}-(t_{max})_1$ se numește fază și exprimă timpul scurs de la momentul maximului precedent și pînă în momentul observației respective. În mod analog $\Phi_{03} = t_{03} - (t_{max})_2$ este faza unei observații din ciclul următor.

Din figura 6.1 se vede că $0 \leqslant \Phi \leqslant P$. Cînd faza ajunge la mărimea unei perioade, se produce trecerea la ciclul următor.

Aşa cum este definită mai sus, faza este exprimată în unităţile în care sînt date abscisele corespunzătoare, dar dacă dorim ca faza să fie exprimată în unităţi de perioadă vom avea :

$$\Phi = \frac{t_{01} - (t_{max})_1}{P} \tag{6.4}$$

astfel că $0 \leqslant \Phi \leqslant 1$.

Noţiunea de fază este deosebit de importantă deoarece, prin intermediul ei, putem reduce toate observaţiile la un singur ciclu. În modul acesta avem posibilitatea să utilizăm şi observaţiile dispersate care ar aparţine la cicluri diferite.

Pentru o observaţie efectuată în ciclul al E-lea putem scrie

$$t = t_E + \Phi = t_0 + E.P + \Phi$$

de unde rezultă faza :

$$\Phi = t - t_0 - E.P.$$

Dacă dorim ca faza să fie exprimată în unităţi de perioadă, putem scrie :

$$\Phi = \frac{t - t_0}{P} - E \tag{6.5}$$

Deoarece $0 \leqslant \Phi \leqslant 1$, din (6.5) se păstrează numai partea fracţionară, numărul E, fiind întreg şi mai mare decît unitatea, trebuie înlăturat prin simpla neluare în considerare.

Observaţie. Pentru utilizarea formulei (6.5) trebuie să cunoaştem precis numărul E, în cazul funcţiilor scurt periodice această condiţie este greu de asigurat din observaţii puţine şi dispersate pe mai multe cicluri situate la distanţe diferite unul de altul.

e) *Exprimarea timpului în zilele iuliene*

În cadrul problemelor de prognoză, utilizarea timpului este de mare importanţă şi aceasta cu atît mai mult cu cît pe noi ne interesează modul de desfăşurare a unor fenomene în viitor. De aceea, considerăm util să prezentăm aici modul de utilizare a timpului exprimat prin intermediul zilelor iuliene.

Utilizarea datei iuliene este foarte simplă, ea se bazează pe alegerea zilei ca unitate de timp, originea de la care se numără aceste zile fiind data de zero ianuarie a anului 4713 dinaintea erei noastre.

Pentru a exprima momentul unei observaţii în zile iuliene, se exprimă momentul respectiv în timp universal care, în ţara noastră, este timpul legal minus 2 ore în timpul iernii şi minus 3 ore vara. Apoi, rezultatul obţinut se exprimă în fracţiuni de zi şi se adună la numărul de zile scurse de la zero ianuarie 4713 î.e.n şi pînă în ziua cînd a fost efectuată observaţia respectivă. În anexa I a volu-

Scanned with OKEN Scanner

mului prezent sînt date zilele iuliene corespunzătoare zilei zero a fiecărei luni pentru orele 12 timp universal.

Pentru motive bine întemeiate, anexa I cuprinde zilele iuliene pentru anii 1980 — 2000. Pentru anii dinaintea lui 1980 se pot utiliza anumite anuare astronomice.

În mod convențional, miezul zilei timp universal a fost ales ca începutul zilei iuliene pentru ca în timpul nopții, cînd se efectuau majoritatea observațiilor astronomice, să nu se schimbe data. Acest reper este arbitrar și nu cauzează nici un neajuns pentru activitatea practică desfășurată în timpul zilei.

Pe baza definiției de mai sus a datei iuliene, momentul unei observații sau măsurători, de exemplu, de la $11^h03^m25^s,3$ timp legal român de la 1 mai 1984 este exprimat prin data iuliană

$$DJ2445821,835709.$$

Acest rezultat se obține în modul următor:

$11^h03^m25^s,3$ este timp legal român și corespunde la $8^h03^m25^s,3$ timp universal.

Știind că $24^h = 1$ zi, $12^h = 0,5$ zile etc, este ușor de calculat fracțiunea de zi care corespunde la $8^h03^m25^s,3$. În acest scop scriem

8^h	0,333333	zile,
3^m	0,002083	,,
25^s	0,000289	,,
$0^s,3$	0,000004	,,
$8^h03^m25^s,3$	0,335709	zile.

Din anexa I, pentru zero mai 1984, orele 12 timp universal, avem data iuliană DJ2445821,0. Pînă la orele 12 timp universal din 1 mai ar fi 1 zi dar observația respectivă s-a efectuat la orele $8^h03^m25^s,3$ timp universal, deci cu ceva înainte de miezul zilei de 1 mai, De aceea, fracțiunea de zi o vom socoti de la miezul zilei de zero mai (= 30 aprilie). Prin urmare, de la orele 12 din zero mai pînă la orele 24 ale aceleiași zile avem $12^h = 0,5$ zile, iar de la orele 24 ale zilei de zero mai sau orele zero ale zilei de 1 mai și pînă la momentul observației au mai trecut $8^h03\ 25^s,3 = 0,335709$ zile. Prin urmare, momentul observației, exprimat în zilele iuliene va fi:

$$\begin{aligned} &DJ2445821,0 \\ &+0,5 \\ &+0,335709 \\ &DJ2445821,835709 \end{aligned}$$

În activitatea practică obișnuită, data iuliană se utilizează în modul indicat mai sus și are avantajul de a simplifica socotirea tim-

pului și utilizarea lui ca variabilă independentă în cadrul cercetării fenomenelor lungi periodice. Chiar și în cazul fenomenelor scurt periodice, utilizarea datei iuliene ușurează mult munca de socotire a timpului deoarece se utilizează o singură unitate de timp — ziua și fracțiunile de zi. De asemenea, în acest mod se simplifică mult calcularea fazelor și întocmirea curbelor medii.

f) *Curba medie a unei funcții empirice periodice*

În paragraful 5.2 am arătat că, atunci cînd avem un număr foarte mare de date empirice, este convenabil a întocmi așa numitele curbe medii care vor avea puncte mai puține, dar mai puțin dispersate. Avantajul acestor curbe constă în aceea că, în cazul funcțiilor periodice, se pot reduce observațiile la un singur ciclu. În acest scop, pentru fiecare observație, cu ajutorul formulei (6.5) se calculează faza corespunzătoare. Apoi, se ordonează punctele după fază și se procedează exact ca în paragraful 5.2, pe porțiuni relativ mici, se aproximează „arcele de curbă" prin segmente liniare și se aplică formulele (5.2).

g) *Determinarea absciselor punctelor de extremum*

În anumite probleme de prognoză, deseori, se ivesc cazuri cînd sîntem interesați în a cunoaște timpul după care anumite fenomene se repetă cu intensitate maximă sau minimă : cutremure puternice de Pămînt, erupții vulcanice cu distrugeri uriașe etc. În astfel de situații am putea să nu ne interesăm de expresia matematică a curbei empirice, am putea să ne preocupăm numai de momentele la care se produc extremele corespunzătoare, adică să determinăm abscisele extremelor funcțiilor empirice.

Pe baza cunoașterii timpului după care se repetă extremele unei funcții, se poate elabora o prognoză destul de bună pentru anumite fenomene periodice, chiar dacă nu se cunoaște modul de comportare a curbelor respective în intervalul dintre astfel de puncte.

De asemenea, cunoașterea punctelor de extremum ale unei funcții empirice ne poate servi la determinarea precisă a duratei perioadei și, prin aceasta, se pot pune în evidență variații regulate sau neregulate ale acesteia. În acest mod se pot descoperi fenomene care se suprapun și dau naștere la curbe empirice foarte complicate al căror aspect se modifică odată cu trecerea timpului.

Avînd în vedere faptul că pe noi ne interesează comportarea unui fenomen în timp, în cele ce urmează ne vom referi mai mult la determinarea absciselor punctelor de extremum. În cazul cînd problema va fi de așa natură încît va trebui să determinăm și ordonata unui astfel de punct, vom utiliza una din metodele de interpolare și vom determina ordonata care corespunde la abscisa respectivă.

Scanned with OKEN Scanner

Pentru determinarea extremelor funcţiilor empirice există mai multe metode (vezi Todoran 1976). Aici însă vom insista numai asupra acelor chestiuni care ne pot fi de folos imediat pentru evaluarea rapidă a absciselor punctelor de extremum. Vom discuta două dintre cele mai operative procedee.

i) *Metoda curbei înjumătăţitoare*

Să presupunem că dintr-o activitate experimentală, legată de modul de manifestare a unui fenomen dat, dispunem de o serie de puncte $M_i(x_i,y_i)$ care sînt situate de o parte şi de alta a unui punct de extremum. Cu alte cuvinte, din observaţii sau măsurători dispunem de o serie de puncte determinate în vecinătatea imediată a unui punct în care un anumit fenomen se manifestă cu intensitate maximă sau minimă. Aceste date la vom utiliza pentru determinarea abscisei punctului de extremum şi, în acest scop, vom proceda în felul următor.

Prin utilizarea unui procedeu de interpolare vom încerca să dispunem, pe ambele ramuri ale curbei respective, de puncte care se află la aceeaşi ordonată. Această condiţie se poate realiza cu uşurinţă dacă punctele de pe o ramură le păstrăm aşa cum au rezultat din activitatea experimentală, iar pentru aceleaşi ordonate (ordonatele punctelor menţinute fixe) să efectuăm interpolarea absciselor corespunzătoare pe cea de a doua ramură. Datele astfel obţinute le reprezentăm grafic împreună cu datele empirice de pe prima ramură. Apoi, unind punctele care au aceeaşi ordonată, dar sînt situate pe ramuri diferite, obţinem o serie de coarde paralele cu axa absciselor. Determinînd punctele de la mijlocul acestor coarde şi trasînd curba care trece prin ele, se obţine aşa numita curbă de înjumătăţire sau *curbă înjumătăţitoare.*

Utilizînd una din metodele expuse în capitolul precedent, trasăm curba empirică printre punctele experimentale. Punctul în care această curbă va fi intersectată de curba înjumătăţitoare se va numi punct de extremum, iar abscisa acestiua este parametrul care ne interesează în problema prognozei.

Metoda curbei înjumătăţitoare este foarte uşor de utilizat în activitatea practică, ea nu face nici o ipoteză asupra formei curbei empirice şi asigură o precizie destul de bună.

În paragraful 5.5 am arătat rolul pe care-l poate juca prezenţa simetriei în cazul unor funcţii empirice, această caracteristică contribuind la simplificarea mai multor probleme. Şi în cazul determinării poziţiei unui punct de extremum, prezenţa simetriei va simplifica mult modul de determinare a abscisei corespunzătoare.

În cazul cînd o curbă empirică prezintă simetrie faţă de o dreaptă care trece prin punctul de extremum şi este paralelă cu axa ordonatelor, curba înjumătăţitoare devine axă de simetrie. Aici, abscisa

Scanned with OKEN Scanner

punctului de extremum coincide cu abscisa medie a punctelor de înjumătățire a coardelor. În modul acesta, de pe un grafic bine executat, se citește abscisa căutată.

ii) *Metoda hîrtiei de calc*

Pentru determinarea extremelor la curbe empirice simetrice se poate utiliza metoda bazată pe reflectarea punctelor observate în jurul axei de simetrie, folosind în acest scop hîrtia de calc. Este de precizat că această metodă nu necesită puncte suplimentare, adică nu face uz de metodele de interpolare.

În practică, pe hîrtie milimetrică se execută reprezentarea grafică a datelor experimentale din vecinătatea unui punct de extremum. În acest scop scara graficului este astfel aleasă încît pantele celor două ramuri să fie de aproximativ 45° respectiv 135°. Apoi, pe hîrtie milimetrică se așază hîrtie transparentă — hîrtie de calc pe care se copiază toate punctele care se văd pe hîrtia milimetrică. De asemenea, cu o linie orizontală se marchează valoarea unei anumite ordonate, iar printr-o linie verticală se notează poziția unei abscise arbitrare din vecinătatea punctului de extremum.

Se întoarce hîrtia de calc cu 180° în jurul unei axe oarecare de simetrie (axă verticală) astfel încît ramura ascendentă să devină ramură descendentă și invers. Hîrtia de calc astfel întoarsă o suprapunem peste hîrtia milimetrică avînd grijă ca linia ordonatei fixe de pe hîrtia de calc să se se suprapună peste linia corespunzătoare de pe hîrtia milimetrică. Apoi, avînd grijă ca axele ordonatelor de pe cele două feluri de hîrtie să fie tot timpul paralele, deplasăm hîrtia de calc (în abscisă) pînă cînd putem considera că am realizat cea mai bună suprapunere a punctelor de pe hîrtia de calc și a celor de pe hîrtia milimetrică. În tot acest timp avem grijă de suprapunerea exactă a liniilor de aceeași ordonată.

După ce se realizează suprapunerea punctelor de pe cele două feluri de hîrtie, citim abscisa de pe hîrtia milimetrică peste care s-a suprapus abscisa fixată pe hîrtia de calc, efectuăm media celor două abscise iar valoarea astfel obținută va fi abscisa extremului respectiv.

6.2. Ameliorarea valorilor perioadelor medii

În general, determinarea periodicităților la funcțiile empirice se efectuează prin aproximații succesive. Cu alte cuvinte, din datele experimentale se evaluează o valoare a perioadei, cum ar fi de exemplu utilizarea formulei (6.2). Apoi, utilizînd un procedeu bazat pe aproximații succesive se caută o valoare medie din ce în ce mai precisă pentru durata perioadei. Vom examina aici două precedee de acest fel.

Scanned with OKEN Scanner

a) *Utilizarea curbei medii pentru determinarea perioadei unei curbe empirice periodice*

Am amintit mai înainte faptul că într-o lucrare anterioară am prezentat mai multe metode privind determinarea periodicităţilor medii (vezi Todoran, 1967). Aici însă vom examina un procedeu care n-a mai fost discutat pînă acum.

Să presupunem că dispunem de o valoare aproximativă a perioadei şi ne servim de ea pentru alcătuirea unei efemeride liniare. În acest caz, folosind relaţia (6.5), calculăm fazele pentru toate datele experimentale disponibile. Apoi transcriem datele respective în ordinea strict crescătoare a fazelor şi întocmim o tabelă de forma celei alăturate, unde y reprezintă ordonata determinată empiric pentru abscisa (momentul) t din care s-a calculat faza Φ.

t	Φ	y
...	...	...
...	...	...
...	...	...

Pe porţiuni relativ mici ale curbei empirice, unde arcul de curbă poate fi aproximat cu un segment liniar, punctele individuale se pot grupa în puncte medii, astfel că

$$\bar{\Phi} = \frac{1}{n}\sum_{i=1}^{n}\Phi_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{6.6}$$

unde n reprezintă numărul de puncte individuale care au fost grupate într-un punct normal (mediu).

Avînd în vedere faptul că prin calcularea fazelor, observaţiile din diferite cicluri au fost aduse la un acelaşi ciclu, această aducere este influienţată de precizia pe care o are perioada utilizată. De aceea, vom considera că pentru cea mai bună valoare a perioadei, diferenţele dintre y observat şi y reprezentat prin dependenţa liniară admisă să fie minime. Cu alte cuvinte, dacă dependenţa liniară este admisă de forma

$$y_i = a_0 + a_1\Phi_i \tag{6.7}$$

atunci, pentru cea mai potrivită durată a perioadei trebuie ca suma pătratelor diferenţelor $y_{obs.} - y_{calc.}$ să fie minimă. Prin urmare,

$$F = \sum_{i=1}^{n}(y_i - a_0 - a_1\Phi_i)^2$$

Scanned with OKEN Scanner

să fie minimă, fapt care ne conduce la sistemul

$$\frac{\partial F}{\partial a_0} = -2\sum_{i=1}^{n}(y_i - a_0 - a_1\Phi_i) = 0$$

$$\frac{\partial F}{\partial a_1} = -2\sum_{i=1}^{n}(y_i - a_0 - a_1\Phi_i)\Phi_i = 0$$

sau

$$\sum_{i=1}^{n}(y_i - a_0 - a_1\Phi_i) = 0$$

$$\sum_{i=1}^{n}(y_i - a_0 - a_1\Phi_i)\Phi_i = 0$$

În ultimele relaţii însumăm şi împărţim la n, astfel că

$$\bar{y} = a_0 + a_1\bar{\Phi}$$

$$\overline{\Phi_i y_i} = a_0\bar{\Phi} + a_1(\overline{\Phi_i^2}) \qquad (6.8)$$

unde am utilizat notaţiile:

$$\overline{\Phi_i y_i} = \frac{1}{n}\sum_{i=1}^{n}\Phi_i y_i \qquad \text{şi} \qquad (\overline{\Phi_i^2}) = \frac{1}{n}\sum_{i=1}^{n}\Phi_i^2$$

Din (6.8) se elimină necunoscutele a_0 şi a_1, astfel că

$$y_c = \frac{\overline{\Phi_i Y_i} - \bar{\Phi}\cdot\bar{Y}}{(\overline{\Phi_i^2}) - (\overline{\Phi})^2}(\Phi - \bar{\Phi}) + \bar{y} \qquad (6.9)$$

Utilizînd formula (6.9), pentru fiecare punct individual de fază Φ_i, care intră într-un punct normal, se obţine o ordonată calculată şi astfel se pot calcula diferenţele $y_i - y_c$ de la toate punctele normale. Cu alte cuvinte

$$S = \sum_{k=1}^{N}\sum_{i=1}^{n}(y_i - y_c)_k^2 \qquad (6.10)$$

unde k reprezintă numărul de ordine al punctelor normale iar N este numărul total de astfel de puncte.

Este uşor de înţeles că y_c, calculat prin intermediul formulei (6.9), va depinde de fazele calculate Φ_i, adică va depinde de durata utilizată a perioadei. De aceea, putem considera că suma (6.10) va avea valoarea minimă atunci cînd pentru perioadă se utilizează durata cea mai apropiată de valoarea reală, în limita erorilor de determinare a datelor empirice.

Scanned with OKEN Scanner

Fie

S_0 suma calculată cu valoarea aproximativă, P_0, a perioadei,

S_+ suma calculată cu valoarea aproximativă a perioadei, dar la care s-a adăugat cantitatea ΔP, și

S_- suma calculată cu valoarea aproximativă a perioadei din care s-a scăzut cantitatea ΔP.

Aici ΔP reprezintă o variație mică care se aplică perioadei pentru a constata tendința de creștere sau descreștere a sumelor S. Se va obține astfel o funcție $S = f(\Delta P)$ pe care, în vecinătatea minimului, o aproximăm prin arcul de parabolă

$$S = a(\Delta P)^2 + b(\Delta P) + c, \tag{6.11}$$

astfel că,

$$\begin{aligned} S_+ &= a(\Delta P)^2 + b(\Delta P) + c \\ S_0 &= \qquad\qquad\quad c \\ S_- &= a(\Delta P)^2 - b(\Delta P) + c \end{aligned}$$

Am obținut astfel un sistem de trei ecuații cu cele trei necunoscute a, b, și c pentru care avem:

$$a = \frac{S_+ + S_- - 2S_0}{2(\Delta P)^2}, \qquad b = \frac{S_+ - S_-}{2(\Delta P)}, \qquad c = S_0 \tag{6.12}$$

Pentru S_{min}, din (6.11) rezultă:

$$\delta_P = -\frac{\Delta P}{2}\,\frac{S_+ - S_-}{S_+ + S_- - 2S_0} \tag{6.13}$$

astfel că valoarea cea mai probabilă a perioadei va fi:

$$P = P_0 - \frac{\Delta P}{2}\,\frac{S_+ - S_-}{S_+ + S_- - 2S_0} \tag{6.14}$$

Aplicație practică

Să considerăm că pe malul unui fluviu, de mare importanță economică, s-a amplasat o stație hidrologică pentru înregistrarea nivelului apei respective și elaborarea unei prognoze privind revărsarea acestuia. Prin intermediul acestei stații s-au obținut foarte multe date numerice de forma $y = f(t)$ unde t este momentul la care s-a înregistrat nivelul apei.

Examinarea datelor empirice amintite mai sus arată că în decurs de 40 de ani (1945 — 1985) au avut loc 6 revărsări care au fost precedate de niște creșteri vertiginoase ale nivelului apei. Aceste creșteri le-am extras în vederea elaborării unei prognoze pentru revărsările din anii următori. Reproducem aceste date în primele două coloane ale tabelei 6.1.

Scanned with OKEN Scanner

Tabela 6.1

t.u.	y	Data iuliană	Δt_0	Φ_0	Φ'_0	$\Delta\Phi''$	y_c	$(y - y_c)^2$
3.03.1945		2431...,						
$9^h\ 36^m$	15	517,900	1467,100	0,9999318	0,318	−645	− 2	289
10 30	54	,938	,138	,9999577	,577	−386	+39	225
10 48	64	,950	,150	,9999659	,659	−304	52	144
10 54	70	,954	,154	,9999686	,686	−277	56	196
11 00	80	,958	,158	,9999714	,714	−249	61	361
11 30	100	,979	,179	,9999857	,857	−106	83	289
11 36	110	,983	,183	,9999884	,884	− 79	88	484
11 48	110	517,992	1467,192	,9999945	0,945	− 18	97	169
12 12	130	518,008	0,008	,0000055	1,055	92	115	225
12 18	138	,013	,013	,0000087	1,087	124	120	324
12 42	154	,029	,029	,0000198	1,198	236	137	289
12 48	160	,033	,033	,0000225	1,225	262	141	361
12 54	162	,038	,038	,0000259	1,259	296	147	265
13 12	180	,050	,050	,0000341	1,341	378	160	400
13 24	190	,058	,058	,0000395	1,395	432	168	484
13 30	195	,063	,063	,0000429	1,429	466	174	441
13 48	208	,075	,075	,0000511	1,511	548	187	441
14 12	230	518,092	0,092	0,0000627	1,627	664	205	625
15.03.1953		2434...,						
19 00	2	452,292	1467,092	0,9999262	0,264	−699	−10	144
19 18	16	,304	,104	,9999346	,346	−617	3	169
19 30	26	,313	,113	,9999407	,407	−556	12	196
19 36	30	,317	,117	,9999434	,434	−529	16	196
19 50	40	,326	,126	,9999496	,496	−467	26	196
20 00	48	,333	,133	,9999543	,543	−420	34	196
20 06	54	,338	,138	,9999577	,577	−386	39	225
20 18	62	,346	,146	,9999632	,632	−331	48	196
20 30	72	,354	,154	,9999686	,686	−277	56	256
20 48	86	,367	,167	,9999775	,775	−188	70	256
21 06	100	,379	,179	,9999857	,857	−106	83	289
21 24	114	,392	1467,192	,9999945	,945	− 18	97	289
21 50	134	,410	0,010	,0000068	1,068	105	117	289
22 24	160	,433	,033	,0000225	1,225	262	141	361
22 30	170	452,438	0,038	0,0000259	1,259	296	147	529
28.03.1961		2437...,						
5 00	15	386,708	1467,108	0,9999373	0,373	−590	7	64
5 06	18	,712	,112	,9999400	,400	−563	11	49
5 12	30	,717	,117	,9999434	,434	−529	16	196
6 00	60	,750	,150	,9999659	,659	−304	52	64
6 30	82	,771	,171	,9999802	,802	−161	75	49
6 36	90	,775	,175	,9999829	,829	−134	79	121
6 42	94	,779	1467,179	,9999857	,857	−106	83	121
7 12	120	,800	0,000	,0000000	1,000	+ 37	106	196
7 30	130	,813	,013	,0000089	1,089	126	120	100
7 42	140	,821	,021	,0000143	1,143	180	128	144
8 12	165	,842	,042	,0000286	1,286	323	151	196
8 30	176	,854	,054	,0000368	1,368	405	164	144
8 36	182	,858	,058	,0000395	1,395	432	168	196
9 00	200	,875	,075	,0000511	1,511	548	187	169
9 30	224	386,896	0,096	0,0000654	1,654	691	209	225

Scanned with OKEN Scanner

Tabela 6.1 (continuare)

t.u.	y	Data iuliană	Δt_0	$_h\Phi_0$	Φ'_0	$\Delta\Phi''$	y_c	$(y-y_c)^2$
9.04.1969		2440...,						
15h 00m	17	321,125	1467,125	0,P9999489	0,489	−474	25	64
15 30	36	,146	,146	,9999632	,632	−331	48	144
15 48	50	,158	,158	,9999714	,714	−249	61	121
16 00	60	,167	,167	,9999775	,775	−188	70	100
16 12	68	,175	,175	,9999829	,829	−134	79	121
16 30	84	,188	1467,188	,9999918	,918	− 45	93	81
16 48	92	,200	0,000	,0000000	1,000	37	106	196
16 52	100	,203	,003	,0000020	1,020	57	109	81
17 00	106	,208	,008	,0000055	1,055	92	115	81
17 12	120	,217	,017	,0000116	1,116	153	124	16
17 30	130	,229	,029	,0000198	1,198	235	137	49
17 48	144	,242	,042	,0000286	1,286	323	151	49
18 00	154	,250	,050	,0000341	1,341	378	160	36
18 06	162	,254	,054	,0000368	1,368	405	164	4
18 30	178	321,271	0,071	,0000484	1,484	521	182	16
22.04.1977		2443...,						
1 00	15	255,542	1467,142	,9999605	0,605	−358	43	784
1 06	18	,546	,146	,9999632	,632	−331	48	900
1 30	40	,563	,163	,9999748	,748	−215	66	676
1 36	42	,567	,167	,9999775	,775	−188	70	784
2 00	60	,583	,183	,9999884	,884	− 79	88	784
2 18	74	,596	1467,196	,9999973	,973	10	102	784
2 30	84	,604	0,004	,0000027	1,027	64	110	676
2 36	90	,608	,008	,0000055	1,055	92	115	625
2 48	98	,617	,017	,0000116	1,116	153	124	676
3 00	108	,625	,025	,0000170	1,170	207	133	625
3 06	114	,629	,029	,0000198	1,198	235	137	529
3 12	118	,633	,033	,0000225	1,225	262	141	529
3 30	130	,646	,046	,0000314	1,314	351	155	625
3 48	146	,658	,058	,0000395	1,395	432	168	484
4 00	154	,667	,067	,0000457	1,457	494	178	576
4 06	160	255,671	0,071	,0000484	1,484	521	182	484
4.05.1985		2446...,						
10 00	5	189,917	1467,117	,9999434	0,434	−529	16	121
10 06	14	,921	,121	,9999461	,461	−502	21	49
10 18	18	,929	,129	,9999516	,516	−447	29	121
10 30	28	,938	,138	,9999577	,577	−386	39	121
10 48	44	,950	,150	,9999659	,659	−304	52	64
11 00	52	,958	,158	,9999714	,714	−249	61	81
11 06	58	,963	,163	,9999747	,747	−216	66	64
11 18	66	,971	,171	,9999803	,803	−160	75	81
11 30	76	,979	,179	,9999857	,857	−106	83	49
11 32	78	,981	,181	,9999870	,870	− 93	85	49
11 48	90	189,992	1467,192	,9999945	,945	− 18	97	49
12 00	98	190,000	0,000	,0000000	1,000	37	106	64
12 24	108	,017	0,017	,0000116	1,116	153	124	256
12 48	136	,033	,033	,0000225	1,225	262	141	25
12 54	142	,038	,038	,0000259	1,259	296	147	25
13 00	146	,042	,042	,0000286	1,286	323	151	25
13 30	170	,063	,063	,0000429	1,429	466	174	16
13 50	186	,076	,076	,0000518	1,518	555	188	4
14 00	194	190,083	0,083	0,0000566	1,566	603	195	1

Scanned with OKEN Scanner

Pentru comoditatea calculelor, data fiecărei observaţii am exprimat-o în zile iuliene, rezultatele fiind cuprinse în coloana a 3-a. Menţionăm că, avînd în vedere intervalul mare de timp, pentru transformarea timpului în zile iuliene ne-am servit de un anuar astronomic, pentru anii 1980 — 2000 putînd fi utilizată anexa I de la sfîrşitul acestui volum. De asemenea, pentru transformarea orelor şi minutelor în fracţiuni de zi există tabele calculate (v. d . ex. Todoran 1976).

Dintr-o reprezentare grafică a funcţiei $y = f(t)$ am evaluat o primă valoare a perioadei medii $P_0 = 1467{,}2$ zile. Prin urmare, luînd ca epocă iniţială data de 3.03.1945, orele 12 timp universal = = 2431518 zile iuliene, moment care corespunde ordonatei $y = 120$, am stabilit efemerida liniară

$$t = 2431518{,}000 + 1467{,}2\,E$$

Cu ajutorul acestei efemeride am calculat momentele pentru care există $y = 120$ înaintea unei serii de date observaţionale. Rezultatele astfel obţinute sînt date în tabela 6.2, unde pentru perioadă am utilizat cele trei valori : P_0, $P_+ = P_0 + \Delta P$ şi $P_- = P_0 - \Delta P$. În ultima coloană a acestei tabele se găseşte numărul întreg de perioade E. În aceste condiţii am utilizat cele trei efemeride liniare

$$(t_c)_0 = 2431518{,}000 + 1467{,}20\,E$$
$$(t_c)_+ = 2431518{,}000 + 1467{,}21\,E$$
$$(t_c)_- = 2431518{,}000 + 1467{,}19\,E$$

În coloana a 4-a a tabelei 6.1 se găseşte diferenţa de timp Δt_0 dintre momentul observaţiei şi momentul teoretic precedent la care a avut loc nivelul apei $y = 120$ (vezi tabela 6.2).

Tab ela 6.2

P_0	$P_0 + \Delta P$	$P_0 - \Delta P$	E
2430050,80	2430050,79	2430050,81	−1
2431518,00	2431518,00	2431518,00	0
2432985,20	2432985,21	2432985,19	1
2434452,40	2434452,42	2434452,38	2
2435919,60	2435919,63	2435919,57	3
2437386,80	2437386,84	2437386,76	4
2438854,00	2438854,05	2438853,95	5
2440321,20	2440321,26	2440321,14	6
2441788,40	2441788,47	2441788,33	7
2443255,60	2443255,68	2443255,52	8
2444722,80	2444722,89	2444722,71	9
2446190,00	2446190,10	2446189,90	10

Scanned with OKEN Scanner

Fazele Φ din coloana a 5-a a tabelei 6.1 se calculează prin intermediul formulei

$$\Phi = \frac{\Delta t}{P}. \tag{6.16}$$

În legătură cu valorile numerice din coloana a 5-a trebuie să precizăm faptul că numărul mare de zecimale de aici este determinat de durata mare a perioadei în raport cu timpul scurt în care are loc creşterea ordonatelor. De aceea, în locul fazei Φ am introdus o nouă variabilă definită prin expresia

$$\Phi' = 10^4\Phi - 9999. \tag{6.17}$$

Avînd în vedere numărul mare de date observaţionale, punctele individuale empirice $M_i(x_i, y_i)$ au fost scrise în ordinea strict crescătoare a fazelor. Apoi s-au alcătuit punctele normale în care primul punct normal conţine numai 10 puncte individuale, al doilea punct normal are 11 puncte individuale etc. În primele două coloane ale tabelei 6.3 sînt date fazele şi ordonatele medii corespunzătoare.

Tabela 6.3

$\bar{\Phi}$	$\bar{y}$	n	$n\bar{y}$	$n\bar{\Phi}$	$n\bar{y}\bar{\Phi}$	$n(\bar{\Phi})^2$
0,387	17	10	170	3,870	65,790	1,497690
,571	35	11	385	6,281	219,835	3,586451
,693	61	9	549	6,237	380,457	4,322241
,792	67	8	536	6,336	424,512	5,018112
0,899	90	12	1080	10,788	970,920	9,698412
1,037	109	10	1090	10,370	1130,330	10,753690
1,173	129	13	1677	15,249	1967,121	17,887077
1,292	155	9	1395	11,628	1802,340	15,023376
1,420	171	10	1710	14,200	2428,200	20,164000
1,565	207	6	1242	9,390	1943,730	14,695350
		98	9834	94,349	11333,235	102,646399

Din primele două coloane ale tabelei 6.3 se vede că punctele normale tind să reprezinte o relaţie liniară de forma :

$$\bar{y} = a_1\bar{\Phi} + a_0.$$

Din această cauză noi nu am aplicat formula (6.9) la punctele individuale din care este alcătuit un punct normal. În cazul prezent am aplicat formula (6.9) la punctele normale din tabela 6.3, ulti-

mele patru coloane ale acestei tabele conținînd calculele necesare pentru acest scop. De aici au rezultat mărimile:

$$\overline{y} = \frac{9834}{98} = 100{,}35, \qquad \overline{\Phi} = \frac{94{,}349}{98} = 0{,}9627449$$

$$\overline{\Phi Y} = 115{,}645225510 \qquad \overline{(n\Phi^2)} = 1{,}0474122347$$

$$\overline{\Phi}\,\overline{Y} = 96{,}611450715 \qquad \overline{(\Phi)^2} = 0{,}9268777425$$

$$19{,}033774795 \qquad 0{,}1205344922$$

Efectuînd înlocuirile valorilor numerice în (6.9), rezultă următoarea formulă:

$$y_c = 0{,}158\ \Delta\Phi'' + 100,$$

prin intermediul căreia am calculat ordonatele y_c din penultima coloană a tabelei 6.1. Mărimile $\Delta\Phi'' = 100\ \Delta\Phi'$ din coloana a 7-a reprezintă diferența dintre faza medie $\overline{\Phi} = 0{,}963$ și faza corespunzătoare. În fine, în ultima coloană a tabelei 6.1 avem pătratele $(y - y_c)^2$ a căror sumă este $S_0 = \Sigma(y - y_c)^2 = 24815$.

Din lipsă de spațiu nu am mai reprodus aici tabelele corespunzătoare pentru P_+ și P_-. Dăm mumai rezultatele:

$$S_+ = 53892, \qquad S_- = 173962.$$

Efectuînd acum înlocuirile în formula (6.14) avem

$$P = 1467{,}20 - \frac{0{,}01}{2} \cdot \frac{53892 - 173962}{53892 + 173962 - 2(24815)}$$

care duce la:

$$P = 1467{,}20337.$$

Observație. În cazul analizat aici s-ar părea că ar fi posibilă o perioadă de două ori mai mare, dar datele experimentale disponibile nu permit acest lucru. Există măsuri care, deși nu au fost utilizate, confirmă ordinul de mărime al perioadei medii determinate.

În urma calculelor efectuate mai sus, admițînd că $P = 1467{,}20337$ este o valoare probabilă a perioadei de revărsare a fluviului amintit mai înainte, pentru prognoza noilor revărsări stabilim următoarea efemeridă, pentru cazurile cînd $y = 120$.

$$t_c = 2431518{,}000 + 1467{,}20337\ E,$$

cu ajutorul căreia calculăm momentele corespunzătoare:

	Dată iuliană	*Dată calendaristică*	*timp universal*	E
$t_c =$	2447657,237	10.05.1989	17^h41	11
	2449124,440	16.05.1993	22 33	12
	2450591,644	23.05.1997	3 28	13

b) *Utilizarea metodei celor mai mici pătrate pentru ameliorarea unei efemeride.*

În cazul cînd, pe un anumit interval de timp, perioada unei funcții empirice poate fi considerată ca o mărime constantă, pentru determinarea duratei medii a acesteia se pot utiliza datele empirice obținute în cicluri diferite. Într-un astfel de caz, așa cum am menționat mai sus, vom obține o durată medie a perioadei, mărime pe care, în cadrul problemelor de prognoză, o putem utiliza în locul perioadei „adevărate".

Dacă dispunem de date experimentale obținute în mai multe cicluri consecutive sau aproape consecutive, și am reușit să determinăm abscisele punctelor de maxim sau de minim, atunci formula (6.3) se poate scrie:

$$(t_E)_{calc.} \tilde{t}_0 + \tilde{P}E \tag{6.18}$$

unde $\tilde{t}_0$ și $\tilde{P}$ sînt valorile aproximative ale epocii inițiale și respectiv perioadei. Aici $(t_E)_{calc.}$ este momentul calculat pentru maximum (sau minimum) din al E-lea ciclu. Același lucru rămîne valabil și dacă ne referim la o anumită abscisă de pe una din ramurile curbei empirice. Expresia (6.18) este o efemeridă aproximativă pe care ne propunem s-o îmbunătățim prin metoda celor mai mici pătrate.

Pe de altă parte, momentul observat $(t_E)_{obs.}$ va fi definit prin formula:

$$(t_E)_{obs.} = \tilde{t}_0 + \Delta t_0 + (\tilde{P} + \Delta P)E \tag{6.19}$$

unde $\tilde{t}_0 + \Delta\tilde{t}_0$ este momentul inițial corect, pe cînd durata corectă a perioadei va fi $\tilde{P} + \Delta P$. Prin urmare, din (6.18) și (6.19) rezultă:

$$O - C = \Delta t_0 + E \cdot \Delta P \tag{6.20}$$

În (6.20) avem corecțiile necunoscute Δt_0 și ΔP pe care trebuie să le adăugăm la valorile aproximative, pentru a obține valorile medii ale epocii inițiale t_0 și a perioadei P.

Dacă scriem ecuația (6.20) pentru fiecare maxim (sau minim) determinat pe cale empirică, obținem un sistem de n ecuații cu două necunoscute $(i = \overline{1, n})$

$$(O-C)_i = \Delta t_0 + \Delta P \cdot \mathbf{E}_i$$

pe care-l rezolvăm prin metoda celor mai mici pătrate și rezultă necunoscutele Δt_0 și ΔP, astfel încît:

$$t_0 = \tilde{t} + \Delta t_0, \qquad P = \tilde{P} + \Delta P. \tag{6.21}$$

Scanned with OKEN Scanner

6.3. Formule empirice pentru prognoza fenomenelor periodice

În paragrafele precedente am vorbit numai despre perioada unei funcții empirice și utilizarea ei în probleme de prognoză, dar, în anumite situații, pe lîngă periodicitate, ne interesează și intensitatea cu care se va manifesta un anumit fenomen. De aceea, ne propunem ca, în paragraful prezent, să analizăm procedeele corespunzătoare stabilirii unor formule empirice periodice.

a) *Utilizarea metodei corecțiilor diferențiale pentru stabilirea formulelor empirice în cazul funcțiilor periodice*

i) Să considerăm că dispunem de o serie de date experimentale x, y, al căror grafic indică existența unei funcții periodice simetrice, asemănătoare cu funcția *sinus* sau *cosinus*. În acest caz, funcția empirică corespunzătoare se poate reprezenta printr-o formulă de forma

$$y = a + b \sin(\omega_0 + \omega_1 \cdot x) \tag{6.22}$$

unde a, b, ω_0 și ω_1 sînt mărimi constante pe care trebuie să le determinăm prin metoda corecțiilor diferențiale (vezi paragraful 4.4). Aici ω_1 se numește frecvență unghiulară și este legată de perioadă prin formula:

$$P = \frac{2\pi}{\omega_1} \tag{6.23}$$

Produsul $\rho = \omega_1 \cdot x$ reprezintă un unghi exprimat în radiani și se numește *fază*, iar ω_0 este *faza inițială* și corespunde la $x = 0$. Coeficientul b este amplitudinea, iar constanta a reprezintă axa de simetrie în ordonată.

În cazul cînd dispunem de un grafic bun al funcției empirice periodice simetrice, de pe desen se evaluează valorile preliminarii: $\tilde{a}$, $\tilde{b}$, $\tilde{\omega}_0$ și $\tilde{\omega}_1$. Apoi, pentru valorile probabile ale constantelor respective vom avea:

$$\left.\begin{aligned} a &= \tilde{a} + \delta a, & b &= \tilde{b} + \delta b, \\ \omega_0 &= \tilde{\omega}_0 + \delta\omega_0, & \omega_1 &= \tilde{\omega}_1 + \delta\omega_1 \end{aligned}\right\} \tag{6.24}$$

Din graficul curbei empirice determinăm abscisele ordonatelor maxime și minime: x_{max} și x_{min}. Apoi, deoarece funcția este simetrică pentru valoarea aproximativă a perioadei scriem:

$$\tilde{P} = 2(x_{max} - x_{min}). \tag{6.25}$$

iar din (6.23) avem:

$$\tilde{\omega}_1 = \frac{2\pi}{\tilde{P}} \tag{6.26}$$

Pentru valoarea aproximativă a constantei a folosim relația:

$$\tilde{a} = \frac{y_{max} + y_{min}}{2} \tag{6.27}$$

Valoarea aproximativă a lui b se determină din considerațiile că

$$y_{max} = a + b \sin 90°$$

sau

$$\tilde{b} = y_{max} - \tilde{a} \tag{6.28}$$

Dispunînd acum de valorile aproximative $\tilde{a}$, $\tilde{b}$, $\tilde{\omega}_0$ și $\tilde{\omega}_1$, din utilizarea formulei (6.22) putem calcula ordonatele y_c, pentru abscisele la care dispunem de ordonate observate. În aceste condiții, din (6.22) avem:

$$(O-C)_i = \delta a + \sin \omega_i \delta b + \tilde{b} \cos \tilde{\omega}_i \delta \omega_0 + \tilde{b} \cdot x_i \cdot \cos \tilde{\omega}_i \cdot \omega \delta_1 \tag{6.29}$$

unde am notat:

$$\omega_i = \omega_0 + \omega_1 \cdot x_i \tag{6.30}$$

În formula (6.29) $O_i = y_i$ sînt ordonatele observate, pe cînd $C_i = (y_c)_i$ reprezintă ordonatele calculate cu formula:

$$(y_c)_i = \tilde{a} + \tilde{b} \sin(\tilde{\omega}_0 + \tilde{\omega}_1 \cdot x_i) \tag{6.31}$$

Aplicație numerică

În primele două coloane ale tabelei 6.4 este dată o curbă empirică periodică, reprezentarea grafică a acesteia fiind reprodusă în fi-

Tabela 6.4

x	y	y_c	$O-C$	$\tilde{\omega}$	$\sin \tilde{\omega}$	$\tilde{b} \sin \tilde{\omega}$	$\cos \tilde{\omega}$	$x \cos \tilde{\omega}$
0	12	12,0	0,0	0°	+0,0000	0,0000	1,0000	0,0000
2	15	14,2	0,8	18	,3090	2,1631	0,9511	1,9022
4	18	16,1	1,9	36	,5878	4,1145	,8090	3,2360
6	18	17,7	0,3	54	,8090	5,6631	,5878	3,5268
8	19	18,7	0,3	72	0,9511	6,6574	,3090	2,4720
10	19	19,0	0,0	90	1,0000	7,0000	,0000	0,0000
12	18	18,7	−0,7	108	0,9511	6,6574	− ,3090	− 3,7080
14	18	17,7	0,3	126	,8090	5,6631	− ,5878	− 8,2292
16	17	16,1	0,9	144	,5878	4,1145	− ,8090	−12,9440
18	14	14,2	−0,2	162	,3090	2,1631	−0,9511	−17,1198
20	13	12,0	1,0	180	,0000	0,0000	−1,0000	−20,0000
22	9	9,8	−0,8	198	− ,3090	−2,1631	−0,9511	−20,9242
24	6	7,9	−1,9	216	− ,5878	−4,1145	− ,8090	−19,4160
26	6	6,3	−0,3	234	− ,8090	−5,6631	− ,5878	−15,2828
28	6	5,3	0,7	252	−0,9511	−6,6574	− ,3090	− 8,6520
30	5	5,0	0,0	270	−1,0000	−7,0000	,0000	0,0000
32	5	5,3	−0,3	288	−0,9511	−6,6574	,3090	9,8880
34	6	6,3	−0,3	306	− ,8090	−5,6631	,5878	19,9852
36	7	7,9	−0,9	324	− ,5878	−4,1145	,8090	29,1240
38	8	9,8	−1,8	342	− ,3090	−2,1613	,9511	36,1418
40	11	12,0	−1,0	360	0,0000	0,0000	1,0000	40,0000

Scanned with OKEN Scanner

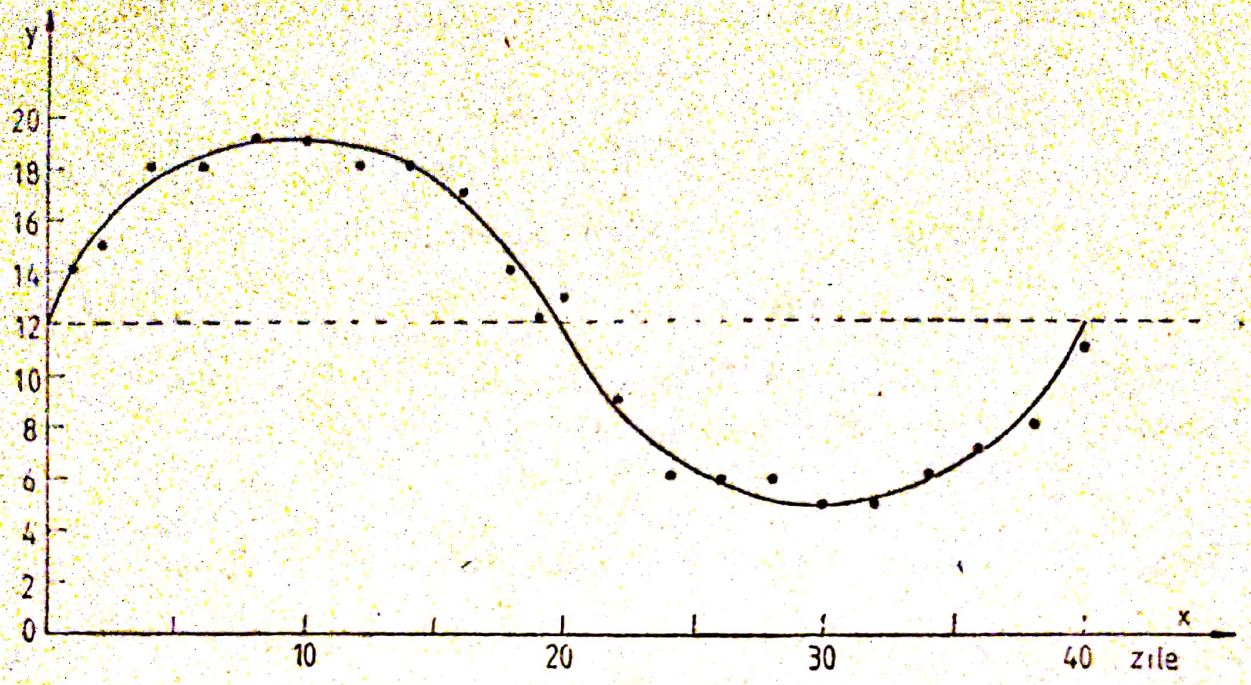


Fig. 6.2. Graficul unei funcţii empirice periodice.

gura 6.2. Din examinarea acestui grafic rezultă abscisa maximului $x_{max} = 10$ iar abscisa minimului este $x_{min} = 30$. Prin urmare, efectuînd înlocuirile în (6.25) obţinem pentru durata medie a perioadei

$$P = 2(30 - 10) = 40 \text{ unităţi de abscisă.}$$

Din (6.26) rezultă :

$$\tilde{\omega}_1 = \frac{360^\circ}{40} = 9 \text{ grade/unitate de abscisă.}$$

Tot de pe graficul din figura 6.2 evaluăm ordonatele aproximative ale punctelor de maxim şi minim :

$$y_{max} = 19, \qquad y_{min} = 5,$$

iar din formula (6.27) rezultă :

$$\tilde{a} = \frac{19 + 5}{2} = 12 \text{ unităţi de ordonată,}$$

Din formula (6.28) avem :

$$\tilde{b} = y_{max} - \tilde{a} = 19 - 12 = 7.$$

Din formula (6.22), pentru $x = 0$, avem

$$\tilde{y}_0 = \tilde{a} + \tilde{b} \sin \omega_0, \qquad \sin \omega_0 = \frac{\tilde{y}_0 - \tilde{a}}{\tilde{b}} = \frac{\tilde{a} - \tilde{a}}{\tilde{b}} = 0$$

adică

$$\tilde{\omega}_0 = 0^\circ.$$

Scanned with OKEN Scanner

Așadar, am determinat valorile aproximative ale celor patru constante din (6.22):

$$\omega_1 = 9°/\text{unitate de abscisă}, \quad \tilde{a} = 12, \quad \tilde{b} = 7 \quad \text{și} \quad \tilde{\omega}_0 = 0°,$$

Cu ajutorul constantelor determinate aici și utilizînd formula (6.31) putem calcula valorile aproximative ale ordonatelor y_c, rezultatele fiind date în cea de a treia coloană a Tabelei 6.4.

Conținutul celorlalte coloane ale tabelei 6.4 este specificat prin capul de coloană corespunzător.

Efectuînd înlocuirile valorilor numerice în formula (6.29) scriem sistemul de ecuații de condiție

$$\begin{aligned}
&X + 0{,}0000\,Y + 1{,}0000\,Z + 0{,}0000\,T = 0{,}0\\
&X + {,}3090\,Y + 0{,}9511\,Z + {,}1902\,T = 0{,}8\\
&X + {,}5878\,Y + {,}8090\,Z + {,}3236\,T = 1{,}9\\
&X + {,}8090\,Y + {,}5878\,Z + {,}3527\,T = 0{,}3\\
&X + 0{,}9511\,Y + {,}3090\,Z + {,}2472\,T = 0{,}3\\
&X + 1{,}0000\,Y + {,}0000\,Z + {,}0000\,T = 0{,}0\\
&X + 0{,}9511\,Y - {,}3090\,Z + {,}3708\,T = -0{,}7\\
&X + {,}8090\,Y - {,}5878\,Z - {,}8229\,T = 0{,}3\\
&X + {,}5878\,Y - {,}8090\,Z - 1{,}2944\,T = 0{,}9\\
&X + {,}3090\,Y - 0{,}9511\,Z - 1{,}7120\,T = -0{,}2\\
&X \quad {,}0000\,Y - 1{,}0000\,Z - 2{,}0000\,T = 1{,}0\\
&X - {,}3090\,Y - 0{,}9511\,Z - 2{,}0924\,T = -0{,}8\\
&X - {,}5878\,Y - {,}8090\,Z - 1{,}9416\,T = -1{,}9\\
&X - {,}8090\,Y - {,}5878\,Z - 1{,}5283\,T = -0{,}3\\
&X - 0{,}9511\,Y - {,}3090\,Z - 0{,}8652\,T = 0{,}7\\
&X - 1{,}0000\,Y - {,}0000\,Z \quad 0{,}0000\,T = 0{,}0\\
&X - 0{,}9511\,Y + {,}3090\,Z + 0{,}9888\,T = -0{,}3\\
&X - 0{,}8090\,Y + {,}5878\,Z + 1{,}9985\,T = -0{,}3\\
&X - 0{,}5878\,Y + {,}8090\,Z + 2{,}9124\,T = -0{,}9\\
&X - {,}3090\,Y + 0{,}9511\,Z + 3{,}6142\,T = -1{,}8\\
&X - 0{,}0000\,Y + 1{,}0000\,Z + 4{,}0000\,T = -1{,}0
\end{aligned}$$

unde am utilizat notațiile:

$$X = \delta a, \qquad Z = \delta'\omega_0 = b.\delta\omega_0$$
$$Y = \delta b, \qquad T = \delta'\omega_1 = 10.b.\delta\Phi_1$$

Prin urmare după aplicarea metodei celor mai mici pătrate am obținut:

$$\delta a = X = -0{,}088 \pm 0{,}176, \qquad \delta b = Y = 0{,}314 \pm 0{,}260$$

$$\delta\omega_0 = \frac{Z}{b} = \frac{0{,}847}{7} = (0{,}121 \pm 0{,}063) \text{ radiani} = 6^g{,}93 \pm 3^g{,}61$$

$$\delta\omega_1 = \frac{T}{70} = \frac{-0{,}4488}{70} = -0^g{,}364 \pm 0{,}153$$

Pentru obținerea rezultatelor căutate înlocuim valorile numerice în expresiile (6.24) și avem

$$a = 11{,}914 \pm 0{,}176, \qquad b = 7{,}314 \pm 0{,}260$$

$\omega_1 = 8^{\circ}{,}638 \pm 0^{\circ}{,}153$ pe unitate de abscisă, $\omega_0 = 6^{\circ}{,}93 \pm 3^{\circ}{,}61$

În aceste condiții, formula empirică corespunzătoare va fi

$$y = 11{,}91 + 7{,}31 \sin (6^{\circ}{,}9 + 8{,}64\, x) \qquad (6.32)$$
$$\pm\, ,18 \pm\ ,26 \quad \pm 3{,}6 \pm 0{,}16$$

ii) În unele cazuri curba empirică se poate reprezenta printr-un polinom trigonometric de forma (2.20) din care să reținem numai

$$y = a_0 + a_1 \cos\omega + a_2 \cos 2\omega + b_1 \sin\omega + b_2 \sin 2\omega \qquad (6.33)$$

unde

$$\omega_i = \frac{360^{\circ}}{P}\, x_i \qquad (6.34)$$

Aici x_i reprezintă abscisele cărora le corespund ordonatele y_i, iar P este perioada exprimată în unități de abscisă.

Considerînd că dispunem de o curbă empirică bine determinată, prin metodele de interpolare, pentru oricare abscisă putem obține ordonata corespunzătoare. Prin urmare, pe graficul curbei empirice alegem punctele caracteristice ale căror ordonate y_0, y_{45}, y_{90}, y_{135}, y_{180}, y_{225}, y_{270} și y_{315} corespund absciselor $\omega = 0^{\circ}$, 45°, 90°, 135°, 180°, 225°, 270° și respectiv 315°. În aceste condiții, din (6.33) avem:

$$\left.\begin{aligned}
y_0 &= a_0 + a_1 + a_2 \\
y_{45} &= a_0 + \frac{\sqrt{2}}{2}\, a_1 + \frac{\sqrt{2}}{2}\, b_1 + b_2 \\
y_{90} &= a_0 - a_2 + b_1 \\
y_{135} &= a_0 - \frac{\sqrt{2}}{2}\, a_1 + \frac{\sqrt{2}}{2}\, b_1 - b_2 \\
y_{180} &= a_0 - a_1 + a_2 \\
y_{225} &= a_0 - \frac{\sqrt{2}}{2}\, a_1 - \frac{\sqrt{2}}{2}\, b_1 + b_2 \\
y_{270} &= a_0 - a_2 - b_1 \\
y_{315} &= a_0 + \frac{\sqrt{2}}{2}\, a_1 - \frac{\sqrt{2}}{2}\, b_1 - b_2
\end{aligned}\right\} \qquad (6.35)$$

Din (6.35) se observă că

$$a_0 = \frac{Y_0 + Y_{45} + Y_{90} + Y_{135} + Y_{180} + Y_{225} + Y_{270} + Y_{315}}{5} \qquad (6.36)$$

Scanned with OKEN Scanner

Din relaţiile 3 şi 7 din (6.35) rezultă

$$b_1 = \frac{Y_{90} - Y_{270}}{2} \tag{6.37}$$

Din relaţiile 1 şi 5 din (6.35) avem

$$\left.\begin{aligned} a_1 &= \frac{Y_0 - Y_{180}}{2} \\ a_2 &= \frac{Y_0 + Y_{180}}{2} - a_0 \end{aligned}\right\} \tag{6.38}$$

Pentru determinarea lui b_2 ne folosim de relaţiile 4 şi 8 din (6.35), astfel că

$$b_2 = a_0 - \frac{Y_{135} + Y_{315}}{2} \tag{6.39}$$

Am determinat aici valorile aproximative $\tilde{a}_i$, $\tilde{b}_i$ ale constantelor din (6.32) şi dispunem de o formulă empirică care se poate utiliza pentru probleme de prognoză.

În cazul cînd ni se cere să determinăm valori foarte precise pentru constantele a_i şi b_i, trebuie să aplicăm metoda corecţiilor diferenţiale care, în cazul formulei (6.33), conduce la un sistem de ecuaţii de condiţie cu 7 necunoscute, numărul ecuaţiilor fiind egal cu numărul de puncte obţinute pe cale experimentală.

b) *Utilizarea analizei armonice pentru stabilirea formulelor empirice la funcţii periodice*

În aplicarea practică a analizei armonice, pentru simplificarea calculelor, se utilizează cazurile cînd se dispune sau se poate dispune de 12, 24, 36 etc ordonate, adică atunci cînd dispunem de un număr de ordonate la care perioada $2\pi = 360°$ se împarte exact. În astfel de cazuri se pot obţine scheme simple pentru calcularea coeficienţilor Fourier.

Cazul cu număr mic de ordonate este uşor de tratat, dar nu se va obţine o formulă empirică destul de precisă, mai ales dacă curba respectivă are şi unele neregularităţi de-a lungul ciclului său.

Cazul cînd se utilizează un număr mare de ordonate va fi mai potrivit, dar este dificil de aplicat. Din aceste motive, în cele ce urmează vom expune numai cazul cînd dispunem de 24 ordonate determinate experimental la abscise echidistante, un caz intermediar între cele două extreme amintite mai sus.

Să considerăm că dispunem de o curbă empirică a cărei perioadă am determinat-o prin una din metodele amintite mai sus. Apoi

intervalul de la zero la 2π îl împărţim în k părţi egale astfel încît ele să fie separate prin abscisele

$$x_1, x_2, \ldots, x_k$$

unde

$$x_\alpha = \frac{2\pi\alpha}{k}, \qquad (\alpha = 1, 2, 3, \ldots, k) \tag{6.40}$$

Prin metodele de interpolare, pentru fiecare x_α determinăm ordonata y_α.

Să considerăm că funcţia empirică periodică se poate exprima prin polinomul trigonometric (2.20) pe care-l scriem sub forma

$$y = f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \tag{6.41}$$

unde, vezi (2.34), avem

$$a_0 = \frac{1}{k} \sum_{\alpha=1}^{k} y_\alpha \tag{6.42}$$

$$a_m = \frac{2}{k} \sum_{\alpha=1}^{k} y_\alpha \cos mx_\alpha, \qquad \left(m \neq 0 \text{ şi } m \neq N\frac{k}{2}\right) \tag{6.43}$$

$$a_m = \frac{1}{k} \sum_{\alpha=1}^{k} y_\alpha \cos mx_\alpha, \qquad \left(m = N\frac{k}{2}\right) \tag{6.44}$$

unde N este un număr întreg.

$$b_m = \frac{2}{k} \sum_{\alpha=1}^{k} y_\alpha \sin mx_\alpha. \tag{6.45}$$

Schema de calcul pentru cazul cînd se cunosc 24 ordonate

Dacă utilizăm 24 de valori ale funcţiei periodice uniform repartizate în intervalul de la 0 la 2π, va trebui să determinăm coeficienţii: $a_0, a_1, a_2, a_3, \ldots, a_{12}$, şi $b_1, b_2, b_3, \ldots, b_{11}$. Prin urmare, $k = 24$, $x_1 = 15°$, $k/2 = 12$.

Din relaţiile (6.42) şi (6.44) avem

$$24a_0 = y_1 + y_2 + y_3 + \ldots + y_{24}, \tag{6.46}$$

$$24a_{12} = -y_1 + y_2 - y_3 + \ldots + y_{24}. \tag{6.47}$$

Scanned with OKEN Scanner

Din relația (6.43), pentru $m = \overline{1,12}$ și $\alpha = \overline{1,24}$, obținem:

$$\left.\begin{aligned}
12a_1 &= y_1 \cos 15^\circ + y_2 \cos 30^\circ + y_3 \cos 45^\circ + y_4 \cos 60^\circ + \\
&+ y_5 \cos 75^\circ + y_6 \cos 90^\circ + y_7 \cos 105^\circ + y_8 \cos 120^\circ + \\
&+ y_9 \cos 135^\circ + y_{10} \cos 150^\circ + y_{11} \cos 165^\circ + y_{12} \cos 180^\circ + \\
&+ y_{13} \cos 195^\circ + y_{14} \cos 210^\circ + y_{15} \cos 225^\circ + y_{16} \cos 240^\circ + \\
&+ y_{17} \cos 255^\circ + y_{18} \cos 270^\circ + y_{19} \cos 285^\circ + y_{20} \cos 300^\circ + \\
&+ y_{21} \cos 315^\circ + y_{22} \cos 330^\circ + y_{23} \cos 345^\circ + y_{24} \cos 360^\circ. \\
12a_2 &= y_1 \cos 30^\circ + y_2 \cos 60^\circ + \ldots + y_{23} \cos 330^\circ + y_{24} \cos 360^\circ. \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
12a_{11} &= y_1 \cos 165^\circ + y_2 \cos 330^\circ + \ldots + y_{23} \cos 195 + y_{24} \cos 360^\circ.
\end{aligned}\right\} \quad (6.48)$$

În mod analog, pentru $m = \overline{1,11}$ și $\alpha = \overline{1,24}$, din (6.45) avem

$$\left.\begin{aligned}
12b_1 &= y_1 \sin 15^\circ + y_2 \sin 30^\circ + y_3 \sin 45^\circ + y_4 \sin 60^\circ + \\
&+ y_5 \sin 75^\circ + y_6 \sin 90^\circ + y_7 \sin 105^\circ + y_8 \sin 120^\circ + \\
&+ y_9 \sin 135^\circ + y_{10} \sin 150^\circ + y_{11} \sin 165^\circ + y_{12} \sin 180^\circ + \\
&+ y_{13} \sin 195^\circ + y_{14} \sin 210^\circ + y_{15} \sin 225^\circ + y_{16} \sin 240^\circ + \\
&+ y_{17} \sin 255^\circ + y_{18} \sin 270^\circ + y_{19} \sin 285 + y_{20} \sin 300^\circ + \\
&+ y_{21} \sin 315^\circ + y_{22} \sin 330^\circ + y_{23} \sin 345^\circ + y_{24} \sin 360^\circ. \\
12b_2 &= y_1 \sin 30^\circ + y_2 \sin 60^\circ + \ldots + y_{23} \sin 330^\circ + y_{24} \sin 360^\circ. \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
12b_{11} &= y_1 \sin 165^\circ + y_2 \sin 330^\circ + \ldots + y_{23} \sin 195^\circ + y_{24} \sin 336^\circ.
\end{aligned}\right\} \quad (6.49)$$

Avînd în vedere formulele (6.43) și (6.45) precum și relațiile complete pentru coeficienții a_1 și b_1, putem scrie cu ușurință toate relațiile pentru coeficienții Fourier,

Din examinarea relațiilor (6.48) și (6.49), rezultă că pentru coeficienții a_m, cosinușii sînt egali și de același semn în rapot cu termenul care-l conține pe y_{12}, adică $\cos 15^\circ = \cos 345^\circ$ etc.

În mod analog, din expresiile coeficienților b_m rezultă că ultimul termen este nul, iar ceilalți se pot grupa doi cîte doi începînd de la $y_1 \sin 15^\circ$ și $y_{23} \sin 345^\circ$ deoarece $\sin 15^\circ = -\sin 345^\circ$.

Pentru gruparea termenilor din (6.46) — (6.49), întocmim Tabela A.

Tabela A

	y_{24}	y_1 y_{23}	y_2 y_{22}	y_3 y_{21}	... ...	y_9 y_{15}	y_{10} y_{14}	y_{11} y_{13}	y_{12}
Sume	s_0	s_1	s_2	s_3	...	s_9	s_{10}	s_{11}	s_{12}
Diferențe		d_1	d_2	d_3	...	d_9	d_{10}	d_{11}	

Scanned with OKEN Scanner

Cu aceste notații, pentru coeficienții Fourier, obținem

$$24a_0 = s_0 + s_1 + s_2 + \ldots + s_{11} + s_{12} \tag{6.50}$$

$$24a_{12} = s_0 - s_1 + s_2 - \ldots - s_{11} + s_{12} \tag{6.51}$$

$$\left.\begin{aligned} 12a_1 &= s_0 + s_1 \cos 15° + s_2 \cos 30° + s_3 \cos 45° + s_4 \cos 60° + \\ &+ s_5 \cos 75° + s_6 \cos 90° + s_7 \cos 105° + s_8 \cos 120° + \\ &+ s_9 \cos 135° + s_{10} \cos 150° + s_{11} \cos 165° + s_{12} \cos 180°. \\ 12a_2 &= s_0 + s_1 \cos 30° + s_2 \cos 60° + \ldots + s_{10} \cos 300° + \\ &+ s_{11} \cos 330° + s_{12} \cos 360°. \end{aligned}\right\} \tag{6.52}$$

$$\left.\begin{aligned} 12b_1 &= d_1 \sin 15° + d_2 \sin 30° + d_3 \sin 45° + d_4 \sin 60° + \\ &+ d_5 \sin 75° + d_6 \sin 90° + d_7 \sin 105° + d_8 \sin 120° + \\ &+ d_9 \sin 135° + d_{10} \sin 150° + d_{11} \sin 165°. \\ 12b_2 &= d_1 \sin 30° + d_2 \sin 60° + \ldots + d_{10} \sin 300° + d_{11} \sin 330°. \end{aligned}\right\} \tag{6.53}$$

Aici putem observa că dacă lăsăm la o parte ultimii termeni cu s_{12}, ceilalți pot fi grupați iarăși doi cîte doi în modul următor:

— pentru m = număr impar, termenii egali depărtați de termenul din mijloc, $a_6 \cos 90°$, vor fi termeni cu cosinuși egali și de semne diferite: $\cos 15° = -\cos 165°$ etc.

— pentru termenii cu m = număr par, cosinușii respectivi vor fi egali și de aceleași semne, de exemplu, $\cos 30° = \cos 330°$. Prin urmare, pentru o nouă restrîngere a termenilor coeficienților a_m vom întocmi tabela B.

Tabela B

s_0 la s_6 s_{12} la s_7	s_0	s_1	s_2	s_3	s_4	s_5	s_6
	s_{12}	s_{11}	s_{10}	s_9	s_8	s_7	
Sumele	c_0	c_1	c_2	c_3	c_4	c_5	c_6
Diferențele	∂_0	∂_1	∂_2	∂_3	∂_4	∂_5	

În cazul formulelor care definesc coeficienții b_m, observăm că pentru m = număr impar, lăsînd la o parte ultimul termen $a_{12} \sin 180°$, ceilalți sînt simetrici în raport cu $s_6 \sin 90°$, de exemplu, $\sin 15° = \sin 165°$. Pentru m = număr impar sinușii respectivi sînt egali și de semne contrare, ex. $\sin 30° = -\sin 330°$.

Scanned with OKEN Scanner

Pe baza ultimelor constatări, pentru restrîngerea termenilor din formulele (6.53) vom întocmi tabela C.

Tabela C

d_1 la d_6 d_{11} la d_7	d_1 d_{11}	d_2 d_{10}	d_3 d_9	d_4 d_8	d_5 d_7	d_6
Sumele	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6
Diferențele	δ_1	δ_2	δ_3	δ_4	δ_5	

Luînd în considerare observațiile și tabelele cu notațiile de mai sus, pentru expresiile coeficienților a_m și b_m obținem:

$$24a_0 = c_0 + c_1 + c_2 + c_3 + c_4 + c_5 + c_6 \tag{6.54}$$

$$24a_{12} = c_0 - c_1 + c_2 - c_3 + c_4 - c_5 + c_6 \tag{6.55}$$

Pentru m = număr impar avem:

$$\left.\begin{aligned}
12a_1 &= \partial_0 + \partial_1 \cos 15° + \partial_2 \cos 30° + \partial_3 \cos 45° + \\
&\quad + \partial_4 \cos 60° + \partial_5 \cos 75° \\
12a_3 &= \partial_0 + \partial_1 \cos 45° + \partial_2 \cos 90° + \partial_3 \cos 135° + \\
&\quad + \partial_4 \cos 180° + \partial_5 \cos 225° \\
12a_5 &= \partial_0 + \partial_1 \cos 75° + \partial_2 \cos 150° + \partial_3 \cos 225° + \\
&\quad + \partial_4 \cos 300° + \partial_5 \cos 15° \\
12a_7 &= \partial_0 + \partial_1 \cos 105° + \partial_2 \cos 210° + \partial_3 \cos 315° + \\
&\quad + \partial_4 \cos 60° + \partial_5 \cos 165° \\
12a_9 &= \partial_0 + \partial_1 \cos 135° + \partial_2 \cos 270 + \partial_3 \cos 45° + \\
&\quad + \partial_4 \cos 180° + \partial_5 \cos 315° \\
12a_{11} &= \partial_0 + \partial_1 \cos 165° + \partial_2 \cos 330° + \partial_3 \cos 135° + \\
&\quad + \partial_4 \cos 300° + \partial_5 \cos 105°
\end{aligned}\right\} \tag{6.56}$$

$$\left.\begin{aligned}
12b_1 &= \sigma_1 \sin 15° + \sigma_2 \sin 30° + \sigma_3 \sin 45° + \sigma_4 \sin 60° + \\
&\quad + \sigma_5 \sin 75° + \sigma_6 \sin 90° \\
12b_3 &= \sigma_1 \sin 45° + \sigma_2 \sin 90° + \sigma_3 \sin 135° + \sigma_4 \sin 180° + \\
&\quad + \sigma_5 \sin 225° + \sigma_6 \sin 270° \\
12b_5 &= \sigma_1 \sin 75° + \sigma_2 \sin 150° + \sigma_3 \sin 225° + \sigma_4 \sin 300° + \\
&\quad + \sigma_5 \sin 15° + \sigma_6 \sin 90° \\
12b_7 &= \sigma_1 \sin 105° + \sigma_2 \sin 210° + \sigma_3 \sin 315° + \sigma_4 \sin 60° + \\
&\quad + \sigma_5 \sin 165° + \sigma_6 \sin 270° \\
12b_9 &= \sigma_1 \sin 135° + \sigma_2 \sin 270° + \sigma_3 \sin 45° + \sigma_4 \sin 180° + \\
&\quad + \sigma_5 \sin 315° + \sigma_6 \sin 90° \\
12b_{11} &= \sigma_1 \sin 165° + \sigma_2 \sin 330° + \sigma_3 \sin 135° + \sigma_4 \sin 300° + \\
&\quad + \sigma_5 \sin 105° + \sigma_6 \sin 270°
\end{aligned}\right\} \tag{6.57}$$

Pentru m = număr par vom putea scrie

$$\left.\begin{aligned}
12a_2 &= c_1 \cos 30^\circ + c_2 \cos 60^\circ + c_3 \cos 90^\circ + c_4 \cos 120^\circ + \\
&+ c_5 \cos 150^\circ + c_6 \cos 180^\circ \\
12a_4 &= c_1 \cos 60^\circ + c_2 \cos 120^\circ + c_3 \cos 180^\circ + c_4 \cos 240^\circ + \\
&+ c_5 \cos 300^\circ + c_6 \cos 360^\circ \\
12a_6 &= c_1 \cos 90^\circ + c_2 \cos 180^\circ + c_3 \cos 270^\circ + c_4 \cos 360^\circ + \\
&+ c_5 \cos 90^\circ + c_6 \cos 180^\circ \\
12a_8 &= c_1 \cos 120^\circ + c_2 \cos 240^\circ + c_3 \cos 360^\circ + c_4 \cos 120^\circ + \\
&+ c_5 \cos 240^\circ + c_6 \cos 360^\circ \\
12a_{10} &= c_1 \cos 150^\circ + c_2 \cos 300^\circ + c_3 \cos 90^\circ + c_4 \cos 240^\circ + \\
&+ c_5 \cos 30^\circ + c_6 \cos 180^\circ
\end{aligned}\right\} \quad (6.58)$$

$$\left.\begin{aligned}
12b_2 &= \delta_1 \sin 30^\circ + \delta_2 \sin 60^\circ + \delta_3 \sin 90^\circ + \delta_4 \sin 120^\circ + \\
&+ \delta_5 \sin 150^\circ \\
12b_4 &= \delta_1 \sin 60^\circ + \delta_2 \sin 120^\circ + \delta_3 \sin 180^\circ + \delta_4 \sin 240^\circ + \\
&+ \delta_5 \sin 300^\circ \\
12b_6 &= \delta_1 \sin 90^\circ + \delta_2 \sin 180^\circ + \delta_3 \sin 270^\circ + \delta_4 \sin 360^\circ + \\
&+ \delta_5 \sin 90^\circ \\
12b_8 &= \delta_1 \sin 120^\circ + \delta_2 \sin 240^\circ + \delta_3 \sin 360^\circ + \delta_4 \sin 120^\circ + \\
&+ \delta_5 \sin 240^\circ \\
12b_{10} &= \delta_1 \sin 150^\circ + \delta_2 \sin 300^\circ + \delta_3 \sin 90^\circ + \delta_4 \sin 240^\circ + \\
&+ \delta_5 \sin 30^\circ
\end{aligned}\right\} \quad (6.59)$$

Relaţiile (6.56) şi (6.57) le lăsăm aşa cum sînt deoarece nu se mai pot efectua restrîngeri avantajoase de termeni care să simplifice calculele în mod esenţial.

În expresiile corespunzătoare lui m = număr par se poate realiza încă o restrîngere. Astfel, pentru coeficienţii a_m unde m este număr par, termenii egal depărtaţi de $c_3 \cos 90^\circ$, $c_3 \cos 180^\circ$ etc, sînt simetrici: $\cos 30^\circ = -\cos 150^\circ$ etc., $\cos 60^\circ = \cos 300^\circ$ etc, $\cos 180^\circ = -\cos 360^\circ$ etc, $\cos 120^\circ = \cos 240^\circ$ etc, $\cos 150^\circ = -\cos 30^\circ$ etc În modul acesta, pentru restrîngerea termenilor din relaţiile (6.58) şi (6.59) vom întocmi tabela D.

Tabela D

c_0 la c_3 c_6 la c_4	c_0 c_6	c_1 c_5	c_2 c_4	c_3
Sumele	l_0	l_1	l_2	l_3
Diferenţele	m_0	m_1	m_2	

Scanned with OKEN Scanner

Utilizînd datele din tabela D, formulele (6.54), (6.55) şi (6.58) devin:

$$24a_0 = l_0 + l_1 + l_2 + l_3 \tag{6.60}$$

$$24a_{12} = l_0 - l_1 + l_2 - l_3 \tag{6.61}$$

$$\left.\begin{aligned} 12a_2 &= m_0 + m_1 \cos 30° + m_2 \cos 60° \\ 12a_4 &= l_0 + l_1 \cos 60° + l_2 \cos 120° + l_3 \cos 180° \\ 12a_6 &= m_0 + m_1 \cos 90° + m_2 \cos 180° \\ 12a_8 &= l_0 + l_1 \cos 120° + l_2 \cos 240° + l_3 \cos 360° \\ 12a_{10} &= m_0 + m_1 \cos 150° + m_2 \cos 300° \end{aligned}\right\} \tag{6.62}$$

Pentru coeficienţii b_m cu m = număr impar, putem face observaţii asemănătoare: sin 30° = sin 150° etc, sin 60° = — sin 240° etc, sin 150° = sin 30° etc. În aceste condiţii întocmim tabela E.

Tabela E

δ_1 la δ_3 δ_5 la δ_4	δ_1 δ_5	δ_2 δ_4	δ_3
Sumele	k_1	k_2	k_3
Diferenţele	n_1	n_2	

Prin urmare, pentru m = număr par avem:

$$\left.\begin{aligned} 12b_2 &= k_1 \sin 30° + k_2 \sin 60° + k_3 \sin 90° \\ 12b_4 &= n_1 \sin 60° + n_2 \sin 120° \\ 12b_6 &= k_1 \sin 90° + k_2 \sin 180° + k_3 \sin 270° \\ 12b_8 &= n_1 \sin 120° + n_2 \sin 240° \\ 12b_{10} &= k_1 \sin 150° + k_2 \sin 300° + k_3 \sin 90° \end{aligned}\right\} \tag{6.63}$$

În formulele stabilite mai sus ar fi posibile şi alte transformări care însă nu ar duce la o modificare esenţială a lor. De aceea, rămîne să ilustrăm modul de aplicare practică a formulelor (6.60)—(6.62) pentru determinarea coeficienţilor a_m pentru m=0 şi numerele pare.

Formulele (6.56) ne permit să determinăm coeficienţii a_m pentru m = număr impar.

Coeficienţii b_m urmează a fi determinaţi din formulele (6.63) şi (6.57).

Rezultatele obţinute pentru a_m şi b_m se înlocuiesc în formula (6.41) şi se obţine funcţia empirică corespunzătoare celor 24 de ordonate care au fost utilizate la întocmirea tabelelor A — B.

Scanned with OKEN Scanner

Aplicație numerică

Pentru a ilustra modul de aplicare practică a analizei armonice în cazul cînd se utilizează 24 de ordonate ($x_1 = 15°$), ne propunem să stabilim formula care să reprezinte funcția empirică periodică din coloanele 2 și 3 ale tabelei 6.5.

Tabela 6.5

α	x_α	y_α	y_o	α	x_α	y_α	y_o
1	15°	37	37,01	13	195	173	172,89
2	30	49	48,98	14	210	180	179,98
3	45	57	56,99	15	225	195	185,01
4	60	74	73,98	16	240	197	195,98
5	75	83	82,99	17	255	188	187,97
6	90	90	90,00	18	270	165	164,98
7	105	106	105,99	19	285	149	149,01
8	120	121	121,04	20	300	138	138,00
9	135	133	132,99	21	315	138	139,01
10	150	147	146,98	22	330	127	126,98
11	165	152	152,09	23	345	51	51,01
12	180	160	159,99	24	360	5	5,38

Pentru determinarea coeficienților a_m și b_m, cu ajutorul datelor din tabela 6.5, întocmim tabelele A — E.

Tabela A

Y_1 la Y_{12}		37	49	57	74	83	90	106	121	133	147	152	160	
Y_{24} la Y_{13}	5	51	127	138	138	149	165	188	197	195	180	173		
Sumele	5	88	176	195	212	232	255	294	318	328	327	325	160	s
Diferențele		−14	−78	−81	−64	−66	−75	−82	−76	−62	−33	−21		d

Tabela B

s_0 la s_6	5	88	176	195	212	232	255	
s_{12} la s_7	160	325	327	328	318	294		
Sumele	165	413	503	523	530	526	255	c
Diferențele	−155	−237	−151	−123	−106	−62		ϑ

Tabela C

d_1 la d_6	−14	−78	−81	−64	−66	−75	
d_{11} la d_7	−21	−33	−62	−76	−82		
Sumele	−35	−111	−143	−140	−148	−75	σ
Diferențele	7	−45	−19	+12	+16		δ

Tabela D

c_0 la c_3	165	413	503	525	
c_8 la c_4	255	526	530		
Sumele	420	939	1033	523	l
Diferențele	−90	−113	−27		m

Tabela E

δ_1 la δ_3	7	−45	−19	
δ_5 la δ_4	16	12		
Sumele	23	−33	−19	k
Diferențele	−9	−57		n

Din tabela D înlocuim valorile lui l în formulele (6.60) și (6.61) și obținem mărimile a_0 și respectiv a_{12}.

Pentru determinarea celorlalți coeficienți a_m, pentru m = număr par, în formulele (6.62) vom înlocui mărimile l și m din tabela D. În acest mod obținem valorile mărimilor a_2, a_4, a_6, a_8, a_{10}.

Pentru determinarea coeficienților a_m, cu m numere impare, folosim formulele (6.56) și valorile numerice $\hat{c}$ din tabela B.

Pentru coeficienți b_m, cu numere pare, folosim formulele (6.63) și tabela E.

Pentru coeficienții b_m, cu m numere impare, folosim formulele (6.57) și valorile σ din tabela C.

Rezultatele astfel obținute sînt date în tabela 6.6.

Tabela 6.6

m	a	b
0	121,46	
1	−56,48	−42,07
2	−16,78	− 3,01
3	− 6,56	− 4,77
4	−12,50	− 4,76
5	− 8,70	+ 1,65
6	− 5,25	+ 3,50
7	− 4,17	+ 3,19
8	− 3,58	+ 3,46
9	− 1,61	+ 1,23
10	− 0,47	+ 1,76
11	+ 0,02	− 0,12
12	− 0,38	

Scanned with OKEN Scanner

Valorile coeficienților a_m și b_m le înlocuim în formula generală (6.41) și rezultă formula empirică căutată

$$\begin{aligned} f(x) = 121{,}46 &- 56{,}48\cos x - 16{,}78\cos 2x - 6{,}56\cos 3x - \\ &- 12{,}50\cos 4x - 8{,}70\cos 5x - 5{,}25\cos 6x - \\ &- 4{,}17\cos 7x - 3{,}58\cos 8x - 1{,}61\cos 9x - \\ &- 0{,}47\cos 10x + 0{,}02\cos 11x - 0{,}38\cos 12x - \\ &- 42{,}07\sin x - 3{,}01\sin 2x - 4{,}77\sin 3x - \\ &- 4{,}76\sin 4x + 1{,}65\sin 5x + 3{,}50\sin 6x + \\ &+ 3{,}19\sin 7x + 3{,}46\sin 8x + 1{,}23\sin 9x + \\ &+ 1{,}76\sin 10x - 0{,}12\sin 11x. \end{aligned}$$

Cu ajutorul acestei formule am calculat mărimile y_c din ultima coloană a tabelei 6.2. Comparația dintre curba observată (y_α) și cea calculată (y_c) arată că formula stabilită pentru cazul celor 24 de ordonate verifică destul de bine datele observaționale și nu-i necesară o reluare cu un număr mai mare de ordonate.

În vederea examinării spectrului funcției empirice din tabela 6.2, ne vom referi la capitolul II de unde, pentru uz practic, formula (2.51) o scriem sub forma

$$\omega_j = \frac{2\pi}{P} j$$

unde j este numărul de ordine al unei armonice. În cazul de față avem $j = \overline{1,12}$. Prin urmare, dacă alegem $\omega_1 = 1$ ca unitate de frecvență, așa cum rezultă din înlocuirea valorilor numerice ($P = 2\pi$), avem imediat frecvențele corespunzătoare pentru celelalte armonice

Dacă ținem seamă de formula (2.50) sau (2.53), putem calcula spectrul amplitudinilor în funcție de frecvență. În acest scop, folosind prima relație din (2.50) și coeficienții Fourier determinați mai sus, rezultă tabela 6.7 unde $R^2(\omega) = a^2(\omega) + b^2(\omega)$. Apoi, făcînd uz de relația (3.143), pentru densitatea spectrală $S_{xx}(\omega)$ avem

$$S_{xx}'(\omega) = \frac{\sqrt{a^2(\omega) + b^2(\omega)}}{P} \qquad (3.143)$$

Valorile densității spectrale sînt date în ultima coloană a tabelei 6.7.

Dacă facem uz de formula (3.141), avem posibilitatea să trecem de la densitatea spectrală $S_{xx}(\omega)$ la funcția de autocorelație $R'_{xx}(\omega)$. În acest scop, considerînd numai partea reală a funcției complexe

$$e^{i\omega\tau} = \cos\omega\tau + i\sin\omega\tau,$$

Tabela 6.7

ω	$a(\omega)$	$b(\omega)$	$c(\omega)$	$S_{xx}(\omega)$
1	−56,48	−42,07	70,43	11,21
2	−16,78	− 3,01	17,05	2,71
3	− 6,56	− 4,77	8,11	1,29
4	−12,50	− 4,76	13,38	2,13
5	− 8,70	+ 1,65	8,86	1,41
6	− 5,25	+ 3,50	6,31	1,00
7	− 4,17	+ 3,19	5,25	0,84
8	− 3,58	+ 3,46	4,98	0,79
9	− 1,61	+ 1,23	2,03	0,32
10	− 0,47	+ 1,76	1,82	0,29
11	+ 0,02	− 0,12	0,12	0,02
12	− 0,38	0,00	0,38	0,06

din (3.141) ne mai rămîne

$$R'_{xx}(\tau) = \frac{1}{2\pi} \int_{\omega_1=1}^{\omega_1=12} S_{xx}(\omega) \cos \omega\tau \cdot d\omega \tag{6.64}$$

unde $\omega = 1, 2, \ldots, 12$, și $\tau = 15°, 30°, \ldots, 345°, 360°$.

Trecînd de la variabila continuă la cea discretă, adică utilizînd integrarea numerică și ținînd seamă de faptul că pasul $h = \Delta\omega = 1$, putem scrie

$$R'_{xx}(\tau) = \frac{1}{2\pi} \sum_{i=1}^{11} S_{xx}(\omega_i) \cos \omega_i\tau, \tag{6.65}$$

de unde, pentru un anumit $\tau =$ constant, obținem mărimea corespunzătoare a funcției de corelație pentru densitatea spectrală. Aici am notat $R'_{xx}(\tau)$ în loc de $R_{xx}(\tau)$ pentru a putea face distincție între cele două feluri de corelații definite de (3.144) și cea de mai sus care se referă la densitatea spectrală.

Deoarece în formulele de mai sus intervine factorul $\cos \omega\tau$, corelația astfel definită va tinde să aibă o variație „periodică", pe cînd funcția de autocorelație definită de relația (3.144) are o tendință de a descrește spre zero cînd τ tinde spre extremitatea dreaptă a intervalului său de definiție.

În tabela 6.8 sînt prezentate calculele efectuate pentru determinarea funcției de autocorelație pentru densitatea spectrală care este dată în ultima coloană a tabelei 6.7. Rezultatele din ultima coloană a tabelei 6.8 indică prezența unei autocorelații puternice, așa cum se vede și din urmărirea valorilor lui $S_{xx}(\omega)$ din ultima coloană a tabelei 6.7. Prezența factorului $\cos \omega\tau$ determină aici o variație simetrică în raport cu $\tau = 180°$.

Tabela 6.8

τ \ ω	1	2	3	4	5	6	7	8	9	10	11	$R_{xx}(\tau)$
15°	+10,83	+2,35	+0,91	+1,06	−0,36	0,00	−0,22	−0,40	−0,23	−0,25	−0,01	+2,178
30	+ 9,71	+1,36	0,00	−1,06	−1,22	−1,00	−0,72	−0,40	0,00	+0,15	+0,01	+1,088
45	+ 7,93	0,00	−0,91	−2,13	−1,00	0,00	+0,59	+0,79	+0,23	0,00	−0,01	+0,874
60	+ 5,60	−1,36	−1,29	−1,06	+0,71	+1,00	+0,42	−0,40	−0,32	−0,15	+0,01	+0,503
75	+ 2,90	−2,35	−0,91	+1,06	+1,36	0,00	−0,81	−0,40	+0,23	+0,25	−0,01	+0,210
90	0,00	−2,71	0,00	+2,13	0,00	−1,00	0,00	+0,79	0,00	−0,29	+0,01	−0,174
105	− 2,90	−2,35	+0,91	+1,06	−1,36	0,00	+0,81	−0,40	−0,23	+0,25	−0,01	−0,669
120	− 5,60	−1,36	+1,29	−1,06	−0,71	+1,00	−0,42	−0,40	+0,32	−0,15	+0,01	−1,131
135	− 7,93	0,00	+0,91	−2,13	+1,00	0,00	−0,59	+0,79	−0,23	0,00	−0,01	−1,301
150	− 9,71	+1,36	0,00	−1,06	+1,22	−1,00	+0,72	−0,40	0,00	+0,15	+0,01	−1,390
165	−10,83	+2,35	−0,91	+1,06	−0,36	0,00	+0,22	−0,40	+0,23	−0,25	−0,01	−1,414
180	−11,21	+2,71	−1,29	+2,13	−1,41	+1,00	−0,83	+0,79	−0,32	+0,29	+0,01	−1,298
195	−10,83	+2,35	−0,91	+1,06	−0,36	0,00	+0,22	−0,40	+0,23	−0,25	−0,01	−1,414
210	− 9,71	+1,36	0,00	−1,06	+1,22	−1,00	+0,72	−0,40	0,00	+0,15	+0,01	−1,390
225	− 7,93	0,00	+0,91	−2,13	+1,00	0,00	−0,59	+0,79	−0,23	0,00	−0,01	−1,301
240	− 5,60	−1,36	+1,29	−1,06	−0,71	+1,00	−0,42	−0,40	+0,32	−0,15	+0,01	−1,131
255	− 2,90	−2,35	+0,91	+1,06	−1,36	0,00	+0,81	−0,40	−0,23	+0,25	−0,01	−0,669
270	0,00	−2,71	0,00	+2,13	0,00	−1,00	0,00	+0,79	0,00	−0,29	+0,01	−0,174
285	+ 2,90	−2,35	−0,91	+1,06	+1,36	0,00	−0,81	−0,40	+0,23	+0,25	−0,01	+0,210
300	+ 5,60	−1,36	−1,29	−1,06	+0,71	+1,00	+0,42	−0,40	−0,32	−0,15	+0,01	+0,503
315	+ 7,93	0,00	−0,91	−2,13	−1,00	0,00	+0,59	+0,79	+0,23	0,00	−0,01	+0,874
330	+ 9,71	+1,36	0,00	−1,06	−1,22	−1,00	−0,72	−0,40	0,00	+0,15	+0,01	+1,088
345	+10,83	+2,35	+0,91	+1,06	−0,36	0,00	−0,22	−0,40	−0,23	−0,25	−0,01	+2,178
360	+11,21	+2,71	+1,29	+2,13	+1,41	+1,00	+0,83	+0,79	+0,32	+0,29	+0,01	+3,502

Scanned with OKEN Scanner

6.4. Prognoza fenomenelor periodice cu perioadă variabilă. Cauzalităţi variabile.

Pînă acum ne-am ocupat de prognoza fenomenelor periodice care depindeau de o perioadă constantă; era cel mai simplu caz în care problemele de prognoză se rezolvă relativ cu destulă uşurinţă. În astfel de cazuri o funcţie liniară de forma (6.3) poate servi ca efemeridă cu ajutorul căreia să se prevadă momentele la care anumite fenomene se vor manifesta cu intensitate maximă sau minimă. Problema devine însă mult mai complicată în cazul cînd durata perioadei respective nu este sau nu poate fi considerată ca o mărime constantă.

Existenţa unor observaţii distribuite uniform pe un interval mare de timp, deseori, scoate în evidenţă faptul că anumite curbe empirice nu se repetă cu o periodicitate strictă şi aceasta cu atît mai mult cu cît între două cicluri există un interval mare de timp. Astfel de situaţii se ivesc atunci cînd însăşi cauza care determină „periodicitatea" respectivă este variabilă, variaţia cauzalităţii reflectîndu-se în modul de variaţie a perioadei. În astfel de cazuri vom avea de-a face cu cauzalităţi variabile cărora le corespund efecte variabile.

Analiza curbelor empirice periodice cu perioadă variabilă va constitui subiectul prezentului paragraf. Aici vom încerca să examinăm felul de abordare a problemelor de prognoză atunci cînd avem de-a face cu fenomene caracterizate prin periodicităţi variabile.

În cazul cînd durata perioadei unei funcţii empirice nu rămîne constantă, însăşi definiţia perioadei trebuie revăzută. În astfel de cazuri, timpul dintre două maxime sau minime succesive şi de acelaşi fel este variabil, iar în definiţia perioadei fenomenului analizat va trebui să includem şi această variaţie. În acest scop, mai întîi vom lucra cu o perioadă aproximativ constantă, vom determina variaţia perioadei, vom da o interpretare fenomenului şi, prin acestea, revenim asupra definiţiei perioadei.

Luarea în considerare a modului de variaţie a unor perioade care definesc anumite funcţii empirice poate duce la descoperirea unor efecte secundare, la punerea în evidenţă a anumitor cauze care perturbă periodicitatea respectivă. Prin urmare, considerarea variaţiei anumitei perioade poate duce la o analiză mult mai amănunţită şi o cunoaştere mult mai profundă a fenomenelor, se poate dispune de o prognoză mult mai demnă de încredere.

Modul în care relaţia (6.41) defineşte o funcţie periodică este determinat de existenţa unei perioade constante. În caz contrar, cînd P = variabil, aspectul graficului funcţiei empirice diferă de la un ciclu la altul. Într-o situaţie de acest fel s-ar părea că avem de-a face cu o funcţie aperiodică, iar prognoza fenomenului respectiv depinde mai mult de iscusinţa şi practica cercetătorului.

Să presupunem că, dintr-o activitate experimentală, dispunem de o serie de date (x_i, y_i) repartizate de-a lungul mai multor cicluri. Într-un astfel de caz putem determina abscisele mai multor puncte de maxim sau minim.

Dacă facem apel la formulele (6.18) — (6.20) și efectuăm o reprezentare grafică a diferențelor $O-C = f(E)$, ne vom putea da seama dacă se pune problema prezenței unei funcții periodice, dar cu perioadă variabilă.

— Dacă diferențele $O-C = f(E)$ reprezintă o dreaptă paralelă cu axa timpului (E) înseamnă că în (6.20) Δt_0 are o valoare remarcabilă, perioada însă este constantă.

— Dacă $O-C = f(E)$ se poate reprezenta printr-o linie dreaptă, corecția perioadei, ΔP, are o mărime constantă și reprezintă panta dreptei respective. Și în acest caz perioada funcției considerate este o mărime constantă.

— Dacă $O-C = f(E)$ are un aspect mai complicat, o parabolă, o sinusoidă, etc, funcția analizată are o perioadă variabilă. Determinarea modului de variație a lui P ne permite să stabilim o nouă funcție empirică $P = P_0 + \Delta P(E)$. Aici P_0 este o mărime constantă față de care se determină variația în timp a perioadei, $\Delta P(E)$.

— Dacă funcția $O-C = f(E)$ are un caracter periodic, acesta poate fi determinat de o variație periodică a perioadei. Astfel, dacă

$$P = P_0 + \Delta P = P_0 + a + b \sin(\omega_0 + \omega_1 E)$$

diferențele $O-C$ se scriu sub forma

$$(O-C)_E = \Delta t_0 + E(\Delta P) = \Delta t_0 + E(P_0 + a) + bE \sin(\omega_0 + \omega_1 E) \quad (6.66)$$

Cercetarea unei astfel de funcții este analogă cu cazul examinat în paragraful precedent.

Și în cazurile amintite aici, problema prognozei revine la analiza unor funcții empirice $O-C = f(E)$, analiză care trebuie efectuată prin una din metodele prezentate anterior. În mod asemănător se abordează problema prognozei și atunci cînd pe ciclul unei curbe empirice apar unele „cocoașe" de aspect și formă variabilă în cursul timpului.

6.5. **Prognoza fenomenelor multiperiodice**

După cum se știe din activitatea practică și cum rezultă și din considerații teoretice, în natură există fenomene al căror mod de manifestare se poate caracteriza prin interferența a două sau mai multe funcții periodice, perioadele acestora putînd fi independente și incomensurabile între ele. În astfel de cazuri problemele de prognoză se complică foarte mult. Este adevărat că există încercări de

a stabili formule empirice care să fie utilizate pentru extrapolare, dar rezultatele nu sînt pe măsura eforturilor.

În cazul cînd din activitatea practică nu dispunem de date suficiente pentru a stabili durata exactă a unui ciclu, putem aplica analiza armonică considerînd că perioada P este egală cu intervalul de timp în care dispunem de date empirice. În astfel de situații, coeficienții Fourier a_m și b_m ne-ar permite să prezentăm datele experimentale într-o formulă care să fie valabilă pe intervalul respectiv, nu în mod obligatoriu și în afara acestui interval. O astfel de formulă ar fi utilă pentru probleme de interpolare, nu însă și pentru o extrapolare sigură. Prin urmare, o formulă de acest fel nu ar fi potrivită pentru problemele de prognoză. Totuși prognoza ne interesează și în astfel de cazuri, mai ales atunci cînd nu dispunem de modelarea matematică corespunzătoare. De aceea, unele probleme de corelație pot fi deosebit de importante aici.

În cazul unei funcții empirice „periodice" am putea determina abscisele și ordonatele punctelor de extremum. Apoi, eventual, s-ar putea stabili o efemeridă liniară de forma (6.3) pentru abscisele extremelor respective, diferențele raportate la aceasta fiind notate prin $O-C$. În fine, s-ar putea căuta o corelație între ordonatele y_k și diferențele $(O-C)_k$ pentru punctele de extremum.

Dacă reușim să dispunem de o efemeridă de forma (6.3) sau una mai complicată, avem deja un mijloc de prognoză pentru momentele la care vor avea loc maximele sau minimele respective: erupții vulcanice, apariția unui cutremur de Pămînt, explozia unei stele etc. În plus, corelația stabilită între amplitudini și diferențele $O-C$ ne permite să prevedem și intensitatea cu care se va manifesta fenomenul examinat. Prin urmare, am putea să dispunem de o efemeridă a extremelor, dar nu de o formulă empirică pe care s-o utilizăm pentru extrapolare. În astfel de situații este necesar să menținem sub observație fenomenul respectiv pînă se constată o repetare a modului său de manifestare, închiderea unui ciclu și începutul altui ciclu putînd fi utilizate pentru determinarea perioadei generale, astfel încît să putem aplica analiza Fourier așa cum a fost expusă în paragraful precedent. Dar dacă avem deja efemerida și corelația amintită mai sus, urmărirea observațională a fenomenului poate fi efectuată după un anumit program, astfel încît în punctele de interes să obținem un număr maxim de date experimentale

În anumite cazuri fenomenele se pot manifesta atît de complicat, încît niciodată să nu se poată determina o perioadă sigură, corelațiile stabilite între anumiți parametri fiind singurele instrumente pentru problemele de prognoză. Un exemplu de acest fel poate fi amintită corelația dintre parametrii fizici ai unei stele și stadiul său de evoluție, de unde rezultă că Soarele nu poate suferi o explozie în următoarele milioane de ani.

Scanned with OKEN Scanner

Numărul de zile de la anul 4713 î.e.n. ziua zero ianuarie, pînă la 12h timp universal (miezul zilei) a zilei de zero din fiecare lună

Anul	Ianuarie 0	Februarie 0	Martie 0	Aprilie 0	Mai 0	Iunie 0	Iulie 0	August 0	Septembrie 0	Octombrie 0	Noiembrie 0	Decembrie 0
1980	244 4239	4270	4299	4330	4360	4391	4421	4452	4483	4513	4544	4574
1981	4605	4636	4664	4695	4725	4756	4786	4817	4848	4878	4909	4939
1982	4970	5001	5029	5060	5090	5121	5151	5182	5213	5243	5274	5304
1983	5335	5366	5394	5425	5455	5486	5516	5547	5578	5608	5639	5669
1984	5700	5731	5760	5791	5821	5852	5882	5913	5944	5974	6005	6035
1985	6066	6097	6125	6156	6186	6217	6247	6278	6309	6339	6370	6400
1986	6431	6462	6490	6521	6551	6582	6612	6643	6674	6704	6735	6765
1987	6796	6827	6855	6886	6916	6947	6977	7008	7039	7069	7100	7130
1988	7161	7192	7221	7252	7282	7313	7343	7374	7405	7435	7466	7496
1989	7527	7558	7586	7617	7647	7678	7708	7739	7770	7800	7831	7861
1990	7892	7923	7951	7982	8012	8043	8073	8104	8135	8165	8196	8226
1991	8257	8288	8316	8347	8377	8408	8438	8469	8500	8530	8561	8591
1992	8622	8653	8682	8713	8743	8774	8804	8835	8866	8896	8927	8957
1993	8988	9019	9047	9078	9108	9139	9169	9200	9231	9261	9292	9322
1994	9353	9384	9412	9443	9473	9504	9534	9565	9596	9626	9657	9687
1995	244 9718	9749	9777	9808	9838	9869	9899	9930	9961	9991	0022*	0052*
1996	245 0083	0114	0143	0174	0204	0235	0265	0296	0327	0357	0388	0418
1997	0449	0480	0508	0539	0569	0600	0630	0661	0692	0722	0753	0783
1998	0814	0845	0873	0904	0934	0965	0995	1026	1057	1087	1118	1148
1999	1179	1210	1238	1269	1299	1330	1360	1391	1422	1452	1483	1513
2000	245 1544	1575	1604	1635	1665	1696	1726	1757	1788	1818	1849	1879

* = Prim[illegible]e trei cifre sînt 245 în loc de 244

Scanned with OKEN Scanner

Ceauşescu Dumitru, Tratarea statistică a datelor chimice analitice, Editura Tehnică 1973.

Cebişev P. L., Izbranie Trudi, Izdatelistvo Akademii Nauk, SSSR, Moskva, 1955, p. 309.

Girlaşu S., Prelucrarea în timp real a semnalelor fizice, Editura Scrisul Românesc, 1978.

Hrenov L. S., Semiznacinie Tabliţi Trigonometriceskih Funcţii, Gasudarstvenoe Izdatelistvo, Moskva—Leningrad, 1951.

Marinescu I., Moineagu C., Niculescu R., Rancu R., Urseanu V., Elemente de statistică matematică şi aplicaţiile ei, Editura Ştiinţifică, 1968.

Mihoc Gh., Bergthaller C., Urseanu V., Procese stochastice. Elemente de teorie şi aplicaţii, Editura Ştiinţifică şi Enciclopedică, 1978.

Moineagu C., Modelarea corelaţiilor în economie, Editura Ştiinţifică, 1974.

Moineagu C., Negură I., Urseanu V., Statistica, Editura Ştiinţifică şi Enciclopedică, 1976.

Oancea E., Rădulescu M., Calculul Probabilităţilor şi Statistică Matematică, Universitatea „Babeş—Bolyai" Cluj-Napoca, 1974.

Odrînţev, V. M., Matematiceskoe opisanie obiectov avtomatizaţii, Izdatelistvo „Maşinostroienie", Moskva, 1965.

Romanovski I. V., Matematiceskaia Statistika, Izdatelistvo Akad. Nauk, Uzbekoi SSR, Taşkent, 1961.

Rumşiski L. Z., Prelucrarea matematică a datelor experimentale, Editura Tehnică, 1974.

Scarborough J. B., Numerical Mathematical Analysis, Fifth Edition, London: Oxford University Press, 1962.

Shchigolov B. M., Matematiceskaia Obrabotka Nablüdenii, Gasudarstvenoe Izdatelistvo Moskva, 1960.

Sileţchi M., Curta L., Modelarea matematică a proceselor sociale, Editura Academiei, RSR, 1972.

Spătaru Al., Teoria Transformării Informaţiei. Semnale şi Perturbaţii, Editura Tehnică, 1965.

Spinrad H., A Study of *VZ Cancri*, în Astrophysical Journal, Vol. 131, p. 134, 1960.

Stancu D. D., Curs şi Culegere de probleme de analiză numerică, Univ. „Babeş—Bolyai", Cluj-Napoca, 1977.

Stanomir D., Stănăşilă O., Metode matematice în teoria semnalelor, Editura Tehnică, 1980.

Stuart R. D., Introducere în Analiza Fourier cu aplicaţii în Tehnică, Editura Tehnică, 1971.

Tolstov G. P., Serii Fourier, Editura Tehnică, 1955.

Todoran I., Tratarea matematică a datelor experimentale. Funcţii empirice. Editura Academiei RSR, 1976.

Whittaker E. T., Robinson G., The Calculus of Observations, Blackie and Son, London Glasgow, Bombay and Toronto, 1926.

Yule G. U., Kendall M. C., Introducere în teoria statisticii, Editura Ştiinţifică, 1969.

Scanned with OKEN Scanner

Redactor: MONICA CREMENE
Tehnoredactor: L. HLAVATHY

Apărut: 1989. Bun de tipar: 24.10.1989. Comanda nr.: 2994.
Coli de tipar: 14. Hîrtie: semivelină 50 g/mp.
Format: 61×86/16.

Tiparul executat sub comanda nr. 15
la ÎNTREPRINDEREA POLIGRAFICĂ CLUJ
Municipiul Cluj-Napoca
B-dul Lenin nr. 146.
Republica Socialistă România

Scanned with OKEN Scanner